374
7.9.70

Psychometrika

VOLUME XXI—1956

JANUARY–DECEMBER

Editorial Council

Chairman:—Harold Gulliksen

Managing Editor:—
Lyle V. Jones

Editors:—Dorothy C. Adkins
Paul Horst

Assistant Managing Editor:—
B. J. Winer

Editorial Board

Dorothy C. Adkins
R. L. Anderson
T. W. Anderson
J. B. Carroll
H. S. Conrad
L. J. Cronbach
E. E. Cureton
Paul S. Dwyer
Allen Edwards
Max D. Engelhart
Wm. K. Estes
Henry E. Garrett
Leo A. Goodman
Bert F. Green
J. P. Guilford
Harold Gulliksen
Paul Horst
Alston S. Householder
Lloyd G. Humphreys
Truman L. Kelley
Albert K. Kurtz
Frederic M. Lord
Irving Lorge
Quinn McNemar
George A. Miller
Wm. G. Mollenkopf
Lincoln E. Moses
George E. Nicholson
M. W. Richardson
R. L. Thorndike
Ledyard Tucker
D. F. Votaw, Jr.

PUBLISHED QUARTERLY

By THE PSYCHOMETRIC SOCIETY

AT 1407 SHERWOOD AVENUE

RICHMOND 5, VIRGINIA

Bureau Edni. Psy. Research
DAVID ... ING COLLEGE
Dated. .....
Acc. No ... .

Bureau: Edi... ...ch
DAVID H... ...E
Dated... 7. 9. 70
Acc. No. J - 374

Psychometrika

CONTENTS

THE ADDITIVE CONSTANT PROBLEM IN MULTI-DIMENSIONAL SCALING 1
SAMUEL J. MESSICK AND ROBERT P. ABELSON

A RAPID NON-PARAMETRIC ESTIMATE OF MULTI-JUDGE RELIABILITY 17
DESMOND S. CARTWRIGHT

A STUDY OF SPEED FACTORS IN TESTS AND ACADEMIC GRADES 31
FREDERIC M. LORD

OPTIMAL TEST LENGTH FOR MAXIMUM DIFFERENTIAL PREDICTION 51
PAUL HORST

THE REGRESSION OF GAINS UPON INITIAL SCORES 67
R. F. GARSIDE

A METHOD OF SCALOGRAM ANALYSIS USING SUMMARY STATISTICS 79
BERT F. GREEN

NOTE ON CARROLL'S ANALYTIC SIMPLE STRUCTURE . . . 89
HENRY F. KAISER

QUINN McNEMAR, *Psychological Statistics, Second Edition* 93
A Review by EDWARD E. CURETON

ROBERT L. THORNDIKE AND ELIZABETH HAGEN, *Measurement and Evaluation in Psychology and Education* 94
A Review by JOHN E. MILHOLLAND

DECISION PROCESSES, Thrall, R. M., Coombs, C. H., and Davis, R. L. (Editors) 96
A Review by R. DUNCAN LUCE

LIST OF MEMBERS OF THE PSYCHOMETRIC SOCIETY . . . 99

VOLUME TWENTY-ONE MARCH 1956 NUMBER 1

Bureau Ednl. Psy. Research
DAVID HARE TRAING COLLEGE
Dated ....
Acc. No. ....

COOPERATIVE GRADUATE SUMMER SESSIONS IN STATISTICS

The University of Florida, North Carolina State College, Virginia Polytechnic Institute, and the Southern Regional Education Board are jointly sponsoring a series of cooperative summer sessions in statistics.

The third of these summer sessions will be held at North Carolina State College, June 11–July 20, 1956. A session is scheduled to be held at Virginia Polytechnic Institute in 1957 and at the University of Florida in 1958. Each summer session lasts six weeks, and each course carries approximately three semester hours of graduate credit.

The 1956 session will be held jointly with the Institute in Quantitative Research Methods in Agricultural Economics, sponsored by the Social Science Research Council. Several statistics courses will be oriented towards economic applications.

The combined faculty for the 1956 summer session and Institute at North Carolina State College will include: Professor R. L. Anderson, North Carolina State College; Professor Gertrude M. Cox, North Carolina State College; Professor David B. Duncan, University of Florida; Professor Alva L. Finkner, North Carolina State College; Dr. Arnold H. E. Grandage, North Carolina State College; Professor Robert J. Hader, North Carolina State College; Assistant Professor Cleon Harrell, North Carolina State College; Professor Earl O. Heady, Iowa State College; Professor Clifford G. Hildreth, Michigan State University; Professor Jack Levine, North Carolina State College; Professor Robert J. Monroe, North Carolina State College; and Assistant Professor Walter L. Smith, University of North Carolina.

Courses to be offered this summer are: Statistical Methods I, Statistical Methods II (Design of Experiments), Statistical Theory I (Probability and Parent Distribution), Statistical Theory II (Sampling Distributions and Inference), Sample Survey Designs, Advanced Analysis II, Advanced Calculus for Statistics, Stochastic Processes, Econometric Methods and Linear Programming. Lectures on Linear Equations (Matrix Algebra) and Production Functions will be given in the Institute program.

Inquiries should be addressed to:

Professor J. A. Rigney
Department of Experimental Statistics
North Carolina State College
Raleigh, North Carolina

Psychometrika

CONTENTS

OPTIMAL TEST LENGTH FOR MAXIMUM ABSOLUTE PREDICTION 111
PAUL HORST AND CHARLOTTE MACEWAN

A LEAST SQUARES SOLUTION FOR PAIRED COMPARISONS WITH INCOMPLETE DATA 125
HAROLD GULLIKSEN

PROPORTIONAL PROFILES AND LATENT STRUCTURE . . 135
W. A. GIBSON

A NON-PARAMETRIC TEST OF CORRELATION USING RANK ORDERS WITHIN SUBGROUPS 145
WARREN S. TORGERSON

MAXIMIZING TEST VALIDITY BY ITEM SELECTION . . . 153
HAROLD WEBSTER

THE VALIDITY OF THE SUCCESSIVE INTERVALS METHOD OF PSYCHOMETRIC SCALING 165
WILLIAM W. ROZEBOOM AND LYLE V. JONES

RELATIONSHIPS BETWEEN TWO SYSTEMS OF FACTOR ANALYSIS 185
C. W. HARRIS

TESTING MESSAGE DIFFUSION IN HARMONIC LOGISTIC CURVES 191
STUART CARTER DODD

A METHOD FOR OBTAINING AN ORDERED METRIC SCALE 207
SIDNEY SIEGEL

ISADORE BLUMEN, MARVIN KOGAN, AND PHILLIP J. MCCARTHY *The Industrial Mobility of Labor as a Probability Process* 217
A Review by WILLIAM FELLER

PAUL E. MEEHL, *Clinical and Statistical Prediction* 218
A Review by ROBERT L. THORNDIKE

VOLUME TWENTY-ONE JUNE 1956 NUMBER 2

Psychometrika

CONTENTS

THE RELATION BETWEEN INFORMATION AND VARIANCE ANALYSES 219
W. R. Garner and William J. McGill

MEASUREMENT OF SUBJECTIVE VALUES 229
Harold Gulliksen

MAXIMIZING TEST BATTERY PREDICTION WHEN THE WEIGHTS ARE REQUIRED TO BE NON-NEGATIVE . 245
Joseph Lev

THE MATCHING PROBLEM 253
Edgar J. Gilbert

A COMPARISON OF GAME THEORY AND LEARNING THEORY 267
Herbert A. Simon

"BEST POSSIBLE" SYSTEMATIC ESTIMATES OF COMMUNALITIES 273
Louis Guttman

RANK-BISERIAL CORRELATION 287
Edward E. Cureton

A NOMOGRAM FOR FACTOR ANALYSIS 291
P. D. Deuel

A NOTE ON THE ESTIMATION OF NONSPRUIOUS CORRELATIONS 295
William H. Angoff

THE VARIANCE OF THE NUMBER OF MUTUAL CHOICES IN SOCIOMETRY 299
Leo Katz and Thurlow R. Wilson

A NOTE ON JENKINS' "IMPROVED METHOD FOR TETRACHORIC r" 305
Joshua A. Fishman

MATHEMATICAL MODELS OF HUMAN BEHAVIOR 307
Review of a Symposium by A. S. Householder

VOLUME TWENTY-ONE SEPTEMBER 1956 NUMBER 3

Psychometrika

CONTENTS

A TRIBUTE TO L. L. THURSTONE 309

THE SCALE GRID: SOME INTERRELATIONS OF DATA MODELS 313

CLYDE H. COOMBS

EFFICIENT ESTIMATION AND LOCAL IDENTIFICATION IN LATENT CLASS ANALYSIS 331

RICHARD B. MCHUGH

THE SELECTION OF JUDGES FOR PREFERENCE TESTING . 349

R. DARRELL BOCK

AN EMPIRICAL EVALUATION OF MULTIDIMENSIONAL SUCCESSIVE INTERVALS 367

SAMUEL J. MESSICK

A NEW SCALING TECHNIQUE FOR ABSOLUTE JUDGMENTS 377

BURTON S. ROSNER

DETERMINATION OF THE NUMBER OF INDEPENDENT PARAMETERS OF A SCORE MATRIX FROM THE EXAMINATION OF RANK ORDERS 383

JOSEPH F. BENNETT

REPORT OF THE TREASURER, PSYCHOMETRIC SOCIETY . 395

REPORT OF THE TREASURER, PSYCHOMETRIC CORPORATION 395

INDEX FOR VOLUME 21 397

VOLUME TWENTY-ONE DECEMBER 1956 NUMBER 4

Erratum in Rozeboom, W. W. and Jones, L. V., The validity of the successive intervals method of psychometric scaling. *Psychometrika*, 1956, **21**, 165–183.

On page 183, last sentence in first paragraph should read:

> "Essentially, what our present analysis has shown is that it is always possible to give a distributional definition to a base scale regardless of whether or not a scale exists which simultaneously normalizes all distributions."

PSYCHOMETRIKA—VOL. 21, NO. 1
MARCH, 1956

THE ADDITIVE CONSTANT PROBLEM IN MULTIDIMENSIONAL SCALING*

SAMUEL J. MESSICK†

EDUCATIONAL TESTING SERVICE

AND

ROBERT P. ABELSON

YALE UNIVERSITY

The problem of choosing the correct additive constant to convert relative interstimulus distances to absolute interstimulus distances in multidimensional scaling is investigated. An artificial numerical example is constructed, and various trial values of the constant are inserted to demonstrate the effect on the multidimensional map of making a variety of incorrect choices. Finally, a general solution to the problem, suggested by Dr. Ledyard R Tucker, is presented; each of the computational steps in this solution is set down for easy reference.

The various procedures that have been thus far described for multidimensional scaling of stimulus objects [Torgerson (7), Attneave (2), Klingberg (4)] all have involved what might be termed the "additive constant problem." A recent variation on Torgerson's procedure [Abelson (1), Messick (5)] also encounters the same problem. All of these procedures set up, at an early stage of the analysis, a matrix S_{jk} of scale values representing distance estimates. Each element, s_{jk}, of this matrix represents the estimated psychological distance between stimulus j and stimulus k. Due to the nature of the scaling procedure used, however, a constant can be added to all the distance estimates without affecting the validity of the scale of psychological distances. This could be expressed by

$$s_{jk} + c = d_{jk} \qquad (j \neq k), \tag{1}$$

where s_{jk} is the relative distance between j and k, and c is the additive constant which re-locates the zero point to produce d_{jk}, the absolute distance between j and k. To put it another way, the distances s_{jk} are all relative; there seems to be no way of directly determining the zero point on the distance scale.

The choice of a particular value of the additive constant will in theory affect the subsequent multidimensional mapping of the stimuli, and thus the investigator is faced with the problem of making the most valid choice, or,

*This study was supported in part by Office of Naval Research Contract N6onr-270-20 and by National Science Foundation Grant G-642 to Princeton University.

†Now at Menninger Foundation.

if you prefer, the best choice according to some pragmatic criterion. The criterion which has been previously suggested is that of minimizing the dimensionality of the stimulus space. Torgerson gave an approximate solution for the additive constant but did not explore the nature of the problem deeply. The following derivation leads to an evaluation of the effect of the constant on the multidimensional structure and provides the groundwork for a general solution to the problem.

The most systematic analytical procedure for arriving at a multidimensional map of the stimuli from the distance matrix D is that given by Torgerson. Each d_{jk} is first squared and these squared distances are arrayed in a matrix. Then a B matrix is calculated from this matrix of squared distances. The multidimensional map is obtained by factoring the B matrix. This procedure can be summarized in the following equations, which embody some simplifications of Torgerson's procedure. For n stimuli the elements b_{jk} of the B matrix can be written as

$$2b_{jk} = \frac{1}{n}\sum_k d_{jk}^2 + \frac{1}{n}\sum_j d_{jk}^2 - d_{jk}^2 - \frac{1}{n^2}\sum_j\sum_k d_{jk}^2 \quad \begin{pmatrix} j = 1, 2, \cdots, n \\ k = 1, 2, \cdots, n \end{pmatrix}. \tag{2}$$

Let the matrix F be the basis for the final mapping. An element f_{jm} of F represents the projection of stimulus j on axis m. The origin of this system is at the centroid of the stimuli.

$$B = FF'. \tag{3}$$

However, the d_{jk} are not available, but the set of scale values s_{jk} can be obtained by any of the suggested scaling methods. s_{jk} is related to d_{jk} for $j \neq k$ by (1). The distance d_{jj} between stimulus j and itself is assumed to be zero; all s_{jj} would ordinarily equal zero as well.

$$d_{jj} = s_{jj} = 0. \tag{4}$$

The constant c is to be added, then, only when $j \neq k$. One equation can express the relationships of (1) and (4).

$$d_{jk} = s_{jk} + c(1 - \delta_k^j), \tag{5}$$

where

$$s_{jj} = 0 \qquad \text{for all } j, \tag{6}$$

and

$$\delta_k^j = \begin{cases} 0 & \text{when} \quad j \neq k \\ 1 & \text{when} \quad j = k. \end{cases} \tag{7}$$

Now the effect of the additive constant on the subsequent analysis can be assessed by substituting (5) into (2) and simplifying with the aid of (7).

Dr. Bert F. Green, Jr., has indicated in a letter to the authors that he has independently arrived at an equivalent derivation. This will show the influence of c on the B matrix, although it will not directly show the influence on the final F. A numerical example to be presented will attempt to assess this influence. The terms of (2) are written

$$d_{jk}^2 = s_{jk}^2 + 2cs_{jk} + c^2(1 - \delta_k^j). \tag{8}$$

Note that $(1 - \delta_k^j)^2 = (1 - \delta_k^j)$. Note further that $(1 - \delta_k^j)s_{jk} = s_{jk}$, since when $j \neq k$, $\delta_k^j = 0$ and $(1 - 0)s_{jk} = s_{jk}$; when $j = k$, $\delta_k^j = 1$ and $(1 - 1)s_{jj} = 0 = s_{jj}$.

$$\frac{1}{n}\sum_k d_{jk}^2 = \frac{1}{n}\sum_k s_{jk}^2 + \frac{2c}{n}\sum_k s_{jk} + \frac{(n-1)}{n}c^2. \tag{9}$$

$$\frac{1}{n}\sum_j d_{jk}^2 = \frac{1}{n}\sum_j s_{jk}^2 + \frac{2c}{n}\sum_j s_{jk} + \frac{(n-1)}{n}c^2. \tag{10}$$

$$\frac{1}{n^2}\sum_j\sum_k d_{jk}^2 = \frac{1}{n^2}\sum_j\sum_k s_{jk}^2 + \frac{2c}{n^2}\sum_j\sum_k s_{jk} + \frac{(n-1)}{n}c^2. \tag{11}$$

Equation (2) can now be written as

$$\begin{aligned} 2b_{jk} = {} & \frac{1}{n}\sum_k s_{jk}^2 + \frac{2c}{n}\sum_k s_{jk} + \frac{(n-1)}{n}c^2 \\ & + \frac{1}{n}\sum_j s_{jk}^2 + \frac{2c}{n}\sum_j s_{jk} + \frac{(n-1)}{n}c^2 \\ & - \frac{1}{n^2}\sum_j\sum_k s_{jk}^2 - \frac{2c}{n^2}\sum_j\sum_k s_{jk} - \frac{(n-1)}{n}c^2 \\ & - s_{jk}^2 - 2cs_{jk} - c^2(1 - \delta_k^j). \end{aligned} \tag{12}$$

Grouping terms according to the power of c yields

$$\begin{aligned} 2b_{jk} = {} & \frac{1}{n}\sum_k s_{jk}^2 + \frac{1}{n}\sum_j s_{jk}^2 - s_{jk}^2 - \frac{1}{n^2}\sum_j\sum_k s_{jk}^2 \\ & + 2c\left(\frac{1}{n}\sum_k s_{jk} + \frac{1}{n}\sum_j s_{jk} - s_{jk} - \frac{1}{n^2}\sum_j\sum_k s_{jk}\right) \\ & + c^2\left(\delta_k^j - \frac{1}{n}\right). \end{aligned} \tag{13}$$

The first four terms give the result if c is set equal to zero and are equivalent to (2) with $s_{jk} = d_{jk}$. Examination of (13) shows that the effect of the additive constant is felt differentially in the diagonal and off-diagonal elements of B. Substitution of typical numbers for the s_{jk} will show that the b_{jk} are often insensitive to various choices of c. The value of (13), however, is that it will serve as a basis for a general solution of the additive constant problem.

A Numerical Example

A numerical example will help in clarifying the role of the additive constant and in assessing its influence on the final scales. Eight points were set up in two dimensions in the shape of a square, with one point placed in the middle. This configuration of nine points is shown in Figure 1. The

TABLE 1

D_{jk}

	A	B	C	D	E	F	G	H	I
A	0	1	2	1	1.414	2.236	2	2.236	2.828
B	1	0	1	1.414	1	1.414	2.236	2	2.236
C	2	1	0	2.236	1.414	1	2.828	2.236	2
D	1	1.414	2.236	0	1	2	1	1.414	2.236
E	1.414	1	1.414	1	0	1	1.414	1	1.414
F	2.236	1.414	1	2	1	0	2.236	1.414	1
G	2	2.236	2.828	1	1.414	2.236	0	1	2
H	2.236	2	2.236	1.414	1	1.414	1	0	1
I	2.828	2.236	2	2.236	1.414	1	2	1	0

Absolute interpoint distances for the two-dimensional configuration of Figure 1.

absolute interpoint distances d_{jk} for this configuration are presented in Table 1. A set of relative interpoint distances can be obtained from Table 1 by utilizing the relationship of (5). The particular set of s_{jk} chosen for this example is found by subtracting one from every element of D_{jk}. This procedure provides the set of relative interpoint distances s_{jk} of Table 2.

In an actual experiment the absolute interpoint distances d_{jk} are not available. The s_{jk} of Table 2 would represent the data and would be obtained from any of the suggested scaling methods. The s_{jk} selected for this problem is the set for which the smallest scale value is placed equal to zero. The problem for the experimenter is to estimate the constant to be added to the s_{jk} (Table 2) in order to obtain a set of d_{jk} (Table 1) which minimizes the dimensionality of the stimulus space. The s_{jk} could not be used directly as d_{jk} by setting $c = 0$ because certain inconsistencies exist among these distances, such as the combination $s_{ab} = 0$, $s_{bc} = 0$, and $s_{ac} = 1$. If $c = 1$ were

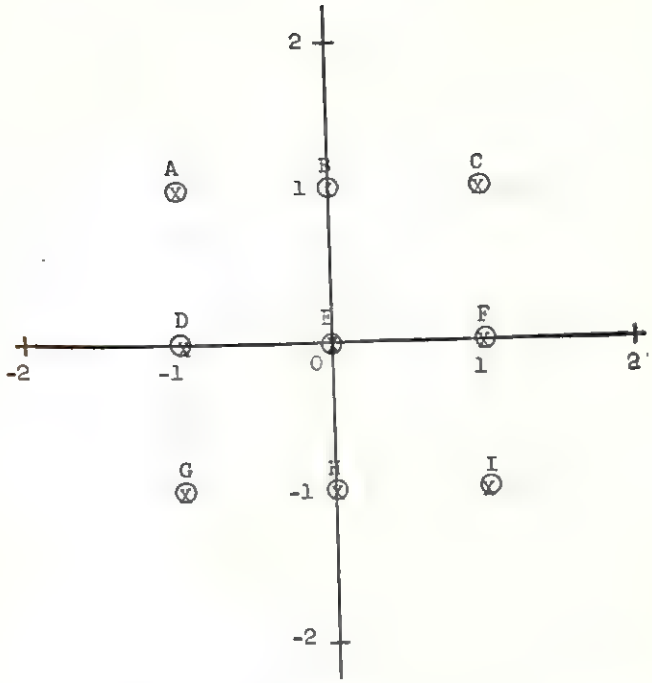


FIGURE 1

Two-dimensional Representation of the Absolute Interpoint Distances, D_{jk}, Given in Table 1

TABLE 2

S_{jk}

	A	B	C	D	E	F	G	H	I
A		0	1	0	.414	1.236	1	1.236	1.828
B	0		0	.414	0	.414	1.236	1	1.236
C	1	0		1.236	.414	0	1.828	1.236	1
D	0	.414	1.236		0	1	0	.414	1.236
E	.414	0	.414	0		0	.414	0	.414
F	1.236	.414	0	1	0		1.236	.414	0
G	1	1.236	1.828	0	.414	1.236		0	1
H	1.236	1	1.236	.414	0	.414	0		0
I	1.828	1.236	1	1.236	.414	0	1	0	

Relative interpoint distances obtained by scaling procedure, setting the smallest distance equal to zero. Additive constant necessary to convert these relative distances into absolute distances, D_{jk}, is c = 1.

TABLE 3

	c = 4		c = 3		c = 2		c = 1		c = 0		c = -1		c = -2	
A	1.91391	1.91391	1.61887	1.61887	1.31631	1.31631	1	1	.65351	.65351	.22056 i	.22056 i	.54627 i	.54627 i
B	2.05751	0	1.72155	0	1.37336	0	1	0	.55035	0	.40210 i	0	.69884 i	0
C	1.91391	-1.91391	1.61887	-1.61887	1.31631	-1.31631	1	-1	.65351	-.65351	.22056 i	-.22056 i	.54627 i	-.54627 i
D	0	2.05751	0	1.72155	0	1.37336	0	1	0	.55035	0	.40210 i	0	.69884 i
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	-2.05751	0	-1.72155	0	-1.37336	0	-1	0	-.55035	0	-.40210 i	0	-.69884 i
G	-1.91391	1.91391	-1.61887	1.61887	-1.31631	1.31631	-1	1	-.65351	.65351	-.22056 i	.22056 i	-.54627 i	.54627 i
H	-2.05751	0	-1.72155	0	-1.37336	0	-1	0	-.55035	0	-.40210 i	0	-.69884 i	0
I	-1.91391	-1.91391	-1.61887	-1.61887	-1.31631	-1.31631	-1	-1	-.65351	-.65351	-.22056 i	-.22056 i	-.54627 i	-.54627 i
β_1*	23.11891		16.41043		10.70289		6		2.31407		-.51797		-2.17044	
β_2*	23.11891		16.41043		10.70289		6		2.31407		-.51797		-2.17044	
β_3	8.01720		4.34520		1.67320		0		-.32860		-.34280		.98520	
β_4	7.32094		3.87824		1.44171		0		-.44191		.11737		1.67946	
β_5	6.98340		3.65540		1.32740		0		-.32860		-.34340		2.01540	
β_6	6.36386		3.24440		1.12391		0		-.14327		.86077		2.68524	
β_7	6.36386		3.24440		1.12391		0		-.14327		.86077		2.68524	
β_8	5.94816		2.96555		.98450		0		.01771		1.03445		3.05274	
β_9	0		0		0		0		0		0		0	

The two principal vectors listed above for each trial value of the additive constant are the vectors which correspond to the configuration of nine points set up for this problem. All nine latent roots of the B matrix for each trial value of c are also listed. The starred values are the roots which correspond to the above vectors.

added to every element of Table 2, the absolute interpoint distances of Table 1 would be obtained; these distances are exactly appropriate to the two-dimensional configuration of Figure 1. No value of the additive constant can be found which would yield a one-dimensional configuration. Thus $c = 1$ is the value which minimizes the dimensionality of the stimulus space. The problem of how to estimate this proper value of the constant, however, remains to be considered.

In an attempt to assess the influence of the additive constant on the configuration of points, a B_{jk} matrix was computed for each of the following trial values of the additive constant: $c = 4, 3, 2, 1, 0, -1, -2$. These seven B_{jk} matrices were obtained from the S_{jk} matrix of Table 2 by substituting the successive values of c into (13). Latent roots and latent vectors were obtained for each B_{jk} (Table 3). All nine roots were extracted for each matrix, and these values were plotted as a function of the additive constant (Figure 2). The "correct" value of c for this S_{jk} is unity, at which point there are two positive roots and seven zero roots. As c changes, the vectors generate a surface, so that for any two values of c, it is possible to identify corresponding vectors. For overestimated values of c (2, 3, 4) the large roots are still easily distinguished from the small ones, but for underestimated values (0, −1, −2) the roots corresponding to the "true" configuration get smaller and finally become negative. Since underestimated values of c offer this possibility of an imaginary space, it is considerably better to overestimate c than to underestimate it. If the configuration is plotted for various values of c (the values for the plots are obtained from Table 3), it can be seen that as c is overestimated more and more, the configuration becomes larger and slightly "convex" (see Figure 3). As c is underestimated more and more the configuration becomes smaller and "concave," but it changes shape at a much faster rate upon underestimation than it does upon overestimation. The configurations for $c = 0$ and $c = 4$ are given in Figure 3. With normalized characteristic vectors, there is smaller relative change when c is large compared to c being small. Since errors arising from the fallibility of scaling procedures would tend to make points which really lie on a straight line deviate from this straight line, it can be shown that the methods of estimating the additive constant described by Torgerson would always give an underestimate.

A General Solution for the Additive Constant

At the "true" solution for the additive constant there is a minimum number of large positive roots and the rest of the roots are zero. With fallible data, however, it should be remembered that the small roots will probably not equal zero but will vary positively and negatively around zero. For any symmetric matrix the sum of all the latent roots is equal to the sum of the diagonal elements. If the sum of only the large latent roots is set equal to the sum of the diagonals, a value of c is obtained which sets the sum of

Bureau Ednl. Psy. Research
ALIANG COLLEGE
Dated
Acc. No

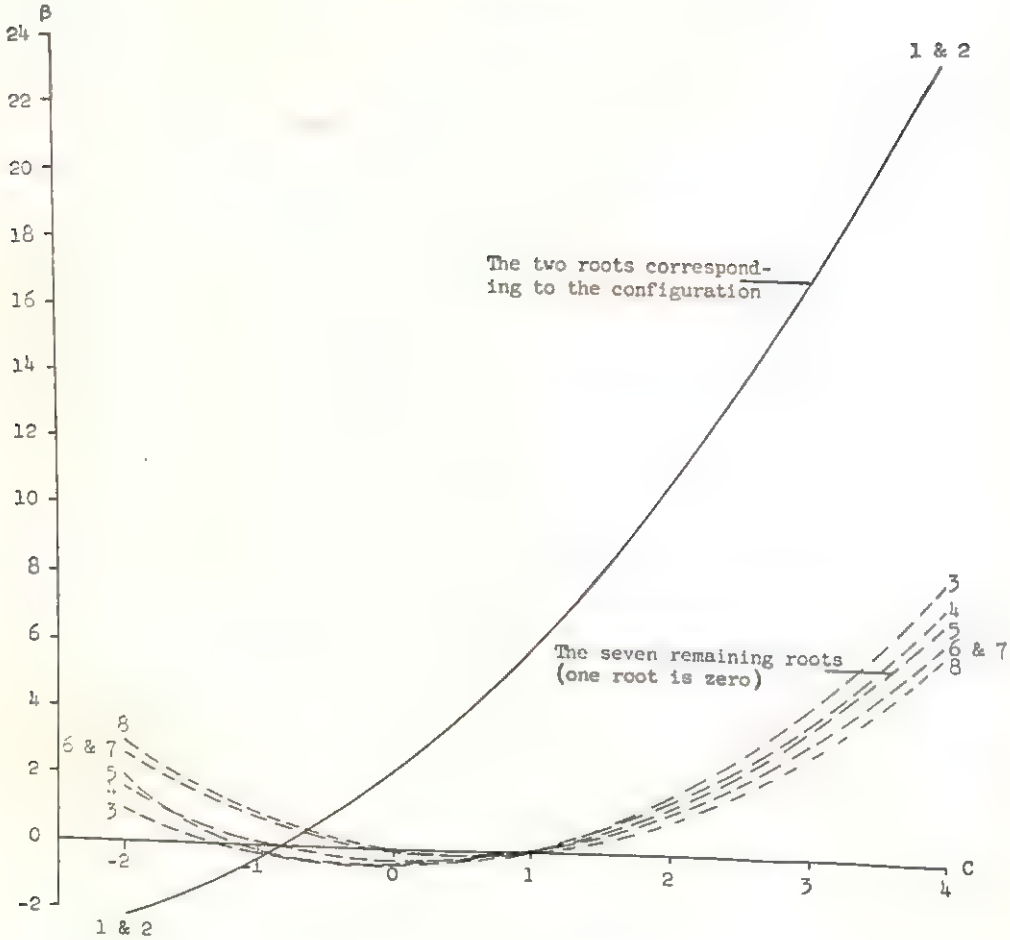


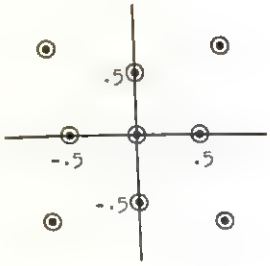
FIGURE 2

The Latent Roots as a Function of Trial Values of the Additive Constant (The value of each root from the proper row of Table 3 plotted against trial values of c. The solution is at c = 1, at which point there are two roots equal to six and seven roots equal to zero.)

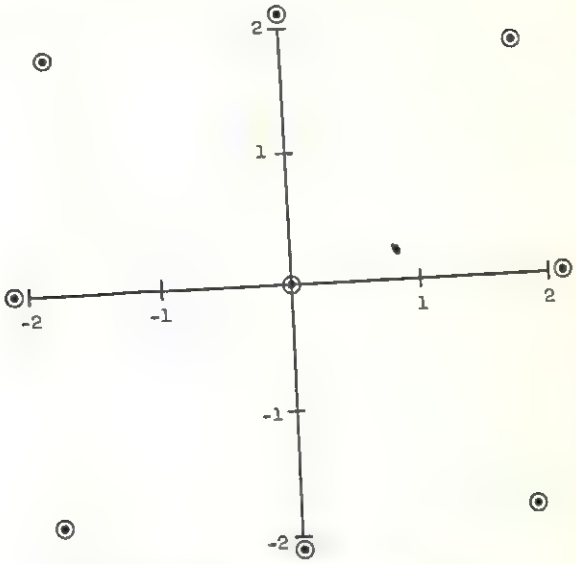
the other roots equal to zero. Dr. Ledyard R Tucker suggested this solution and has made this general approach practical by contributing much in the way of matrix simplification.

From (13) the diagonal elements of B_{ik} may be written as

$$b_{ii} = \frac{1}{2}\left(\frac{1}{n}\sum_k s_{ik}^2 + \frac{1}{n}\sum_i s_{ik}^2 - \frac{1}{n^2}\sum_i\sum_k s_{ik}^2\right)$$
$$+ c\left(\frac{1}{n}\sum_k s_{ik} + \frac{1}{n}\sum_i s_{ik} - \frac{1}{n^2}\sum_i\sum_k s_{ik}\right) \qquad (14)$$
$$+ \frac{1}{2}c^2\left(1 - \frac{1}{n}\right).$$



Configuration of Points Corresponding to c = 0



Configuration of Points Corresponding to c = 4

FIGURE 3

Note that for $j = k$, $\sum_k s_{jk}^2 = \sum_j s_{jk}^2$ and $\sum_k s_{jk} = \sum_j s_{jk}$.

$$b_{jj} = \frac{1}{n}\sum_k s_{jk}^2 - \frac{1}{2n^2}\sum_j\sum_k s_{jk}^2 + c\left(\frac{2}{n}\sum_k s_{jk} - \frac{1}{n^2}\sum_j\sum_k s_{jk}\right) + \frac{1}{2}c^2\left(1 - \frac{1}{n}\right). \tag{15}$$

The sum of the diagonal elements is

$$\sum_j b_{jj} = \frac{1}{n}\sum_j\sum_k s_{jk}^2 - \frac{1}{2n}\sum_j\sum_k s_{jk}^2 + c\left(\frac{2}{n}\sum_j\sum_k s_{jk} - \frac{1}{n}\sum_j\sum_k s_{jk}\right) + \frac{n}{2}c^2\left(1 - \frac{1}{n}\right). \tag{16}$$

Simplifying,

$$\sum_j b_{jj} = \frac{1}{2n}\sum_j\sum_k s_{jk}^2 + \frac{c}{n}\sum_j\sum_k s_{jk} + \frac{1}{2}(n-1)c^2. \tag{17}$$

Let X_1 be the latent vector corresponding to the first latent root β_1.

$$BX_1 = \beta_1 X_1 . \tag{18}$$

If X_1 is normalized so that the sum of the squares of its elements is equal to unity, then,

$$X_1'BX_1 = \beta_1 . \tag{19}$$

The sum of the large latent roots is

$$\sum \beta = X_1'BX_1 + X_2'BX_2 + \cdots + X_p'BX_p\,, \tag{20}$$

where p is the number of large principal components.

Equation 13 may be rewritten in matrix notation as

$$B = A + cE + \tfrac{1}{2}c^2H, \tag{21}$$

where the elements a_{jk}, e_{jk}, and h_{jk} of matrices A, E, and H, respectively, are

$$a_{jk} = \frac{1}{2}\left(\frac{1}{n}\sum_k s_{jk}^2 + \frac{1}{n}\sum_j s_{jk}^2 - s_{jk}^2 - \frac{1}{n^2}\sum_j\sum_k s_{jk}^2\right), \tag{22}$$

$$e_{jk} = \frac{1}{n}\sum_k s_{jk} + \frac{1}{n}\sum_j s_{jk} - s_{jk} - \frac{1}{n^2}\sum_j\sum_k s_{jk}\,, \tag{23}$$

$$h_{jk} = -\frac{1}{n}\quad (j \neq k), \qquad \text{and} \qquad h_{jj} = \left(1 - \frac{1}{n}\right). \tag{24}$$

Substituting (21) in (20),

$$\begin{aligned}\sum \beta = {} & (X_1'AX_1 + cX_1'EX_1 + \tfrac{1}{2}c^2X_1'HX_1)\\ & + (X_2'AX_2 + cX_2'EX_2 + \tfrac{1}{2}c^2X_2'HX_2)\\ & + \cdots + (X_p'AX_p + cX_p'EX_p + \tfrac{1}{2}c^2X_p'HX_p).\end{aligned} \tag{25}$$

$$\sum \beta = \sum_i^p X_i'AX_i + c\sum_i^p X_i'EX_i + \frac{1}{2}c^2\sum_i^p X_i'HX_i\,. \tag{26}$$

In the following consideration of the matrix H, it is found that the last term of (26) can be simplified. The diagonal elements of H are $(1 - 1/n)$ and the off-diagonal elements are $(-1/n)$. Since the sum of each row or column is equal to zero, one of the roots of the H matrix is zero. The other $(n - 1)$ roots are all equal to unity. The vector corresponding to the zero root has equal coefficients, and the other $(n - 1)$ vectors are indeterminate. However, any vector which has a sum of coefficients equal to zero is a possible vector for H. Since the sum of each row or column of B is also equal to zero (it is a centroid matrix), every B matrix also has a zero root for which the vector has equal coefficients. In order to maintain the orthogonality of the principal vectors of B, the sum of the coefficients of each of the other vectors must equal zero. Therefore, the vectors of B are possible vectors of H. The vector with the vanishing root is not considered in either B or H.

$$X_1'HX_1 = \beta_{H1} = 1. \tag{27}$$

$$\sum_i^p X_i'HX_i = \sum_i^p \beta_{Hi} = p \qquad (p = \text{number of roots taken}). \tag{28}$$

Therefore, (26) can be written

$$\sum \beta = \sum_{i}^{p} X_i' A X_i + c \sum_{i}^{p} X_i' E X_i + \frac{1}{2} p c^2. \tag{29}$$

If the sum of the large latent roots of (29) is set equal to the sum of the diagonals of (17) to yield a quadratic in c, the solution of this quadratic equation will give the value of c which sets the sum of the small latent roots equal to zero.

The solution outlined in (17) to (29) involves several computational steps. In any of the suggested scaling procedures it is necessary to obtain a matrix of relative interpoint distances S_{jk} which can be adjusted so that the smallest interpoint distance is set equal to zero. A trial value of c must then be selected, and it is much better to overestimate than to underestimate. In practice either the largest value or the average value in the S_{jk} matrix could be used as a trial value. If some of the stimuli are relatively close together these will usually provide an overestimate of c. In any event, the value of the additive constant obtained in this solution will indicate whether c was overestimated and is being approached from above, or whether it was underestimated and is being approached from below.

A Numerical Illustration of the General Solution for the Additive Constant

In order to try out the solution, the "data" of Table 2 were selected, and a trial value of $c = 4$ was chosen. This trial value was then inserted into (13) to obtain a B matrix (Table 4). A principal components analysis of B yields the X_i vectors (the two largest principal components for $c = 4$ are given in the first two columns of Table 3). It is not necessary, however, to

TABLE 4

B_{jk}

	A	B	C	D	E	F	G	H	I
A	11.44940	2.14480	-1.05060	2.14480	-.97830	-3.56305	-1.05060	-3.56305	-5.53340
B	2.14480	8.84020	2.14480	-.90150	-.54090	-.90150	-3.56305	-3.65980	-3.56305
C	-1.05060	2.14480	11.44940	-3.56305	-.97830	2.14480	-5.53340	-3.56305	-1.05060
D	2.14480	-.90150	-3.56305	8.84020	-.54090	-3.65980	2.14480	-.90150	-3.56305
E	-.97830	-.54090	-.97830	-.54090	6.07760	-.54090	-.97830	-.54090	-.97830
F	-3.56305	-.90150	2.14480	-3.65980	-.54090	8.84020	-3.56305	-.90150	2.14480
G	-1.05060	-3.56305	-5.53340	2.14480	-.97825	-3.56305	11.44940	2.14480	-1.05060
H	-3.56305	-3.65980	-3.56305	-.90150	-.54090	-.90150	2.14480	8.84020	2.14480
I	-5.53340	-3.56305	-1.05060	-3.56305	-.97830	2.14480	-1.05060	2.14480	11.44940

B_{jk} matrix based upon an additive constant of $c = 4$.

extract all of the principal components—only the large ones. Inasmuch as the centroid method is an approximation to principal axes, a centroid analysis could be used if the number of factors extracted is not large. Iterations of this solution will then yield closer and closer approximations to principal components as well as closer approximations to c. The X_i vectors must then be normalized.

$$X_1 \text{ (normalized)} = \begin{bmatrix} .39805 \\ .42792 \\ .39805 \\ 0 \\ 0 \\ 0 \\ -.39805 \\ -.42792 \\ -.39805 \end{bmatrix}; \qquad X_2 \text{ (normalized)} = \begin{bmatrix} .39805 \\ 0 \\ -.39805 \\ .42792 \\ 0 \\ -.42792 \\ .39805 \\ 0 \\ -.39805 \end{bmatrix}.$$

The values of $\sum_i \sum_k s_{ik}^2$ and $\sum_i \sum_k s_{ik}$ obtained from Table 2 are then substituted in (17) to yield the sum of the diagonals of B.

$$\sum_i b_{ii} = 2.91956 + 5.07911c + 4c^2. \tag{30}$$

Matrices A and E are constructed using (22) and (23). Substituting the appropriate numerical values into (29) gives

$$\sum \beta = 4.56580 + 6.41804c + c^2. \tag{31}$$

Equating (30) and (31) and solving, $c = .997$ and $c = -.55$.

The graph of Figure 2 indicates that it is considerably easier to distinguish the large roots from the small ones when the sum of the latent roots is large. The value of c desired, then, is the one which will give the highest $\sum \beta$. Substituting the two roots into (29), the desired c is seen to be .997.

$$\sum \beta = 4.56580 + 6.41804\,(.997) + (.997)^2 = 11.96$$

$$\sum \beta = 4.56580 + 6.41804\,(-.55) + (-.55)^2 = 1.34.$$

The correct value of the additive constant for the S_{ik} of Table 2 is unity. The solution of .997 may be considered only an approximation to this correct value, since the X_i vectors are obtained from a B matrix based on a trial value of c. But this approximation is probably adequate without recourse to an iterative procedure, since the latent vectors do not differ widely for different *overestimated* trial values of c (see Table 3), and the coefficients of the other equation involved in the solution (17) are independent of c.

If the obtained c is satisfactory, it can be used with (13) to produce another B matrix, which can then be factored to yield the final F. If a more exact solution for c were desired, the columns of F could be substituted for the X_i vectors in (29) to obtain a new equation for $\sum \beta$ as a function of c. This new expression for $\sum \beta$ could then be set equal to the expression for the sum of diagonals of B^* (17), and a closer approximation to c would be obtained. If it became apparent that the original trial value was an underestimate, it might be necessary to re-calculate the X_i vectors with a new trial c, because the underestimated value might have led to the choice of the wrong dimensions or to the addition of extraneous dimensions (see Figure 2).

It is also possible, however, to obtain the F matrix without computing another B, by using the general matrix factoring solution (3), which is summarized in the following two equations:

$$(BX)K^{-1} = F, \tag{32}$$

where X is a matrix of the X_i vectors, and K is a matrix for which

$$X'BX = K'K. \tag{33}$$

The matrix $(X'BX)$ is symmetric and should be factored, say by the diagonal or square root method (6), to obtain K'. K^{-1} can then be computed and applied in (32). It was seen above that

$$B = A + cE + \tfrac{1}{2}c^2H, \tag{21}$$

$$BX = (AX) + c(EX) + \tfrac{1}{2}c^2(HX). \tag{34}$$

Thus, (BX) can be found without computing B. It should be pointed out that (AX), (EX), and (HX) are already available from the computations involved in (29). In practice, however, it might be desirable to compute B anyway, in order to evaluate residuals.

Before summarizing this procedure, it is important to point out that the additive constant problem should be regarded in the light of the purpose for which a multidimensional scale is being constructed. The investigator might, on the one hand, wish to do a broad exploratory study [cf. (1)], or, on the other hand, he may be interested in a fairly sensitive analysis [cf. (8)]. In the latter case, a slight dislocation of the stimuli in the psychological space might be damaging, and it would be important to get a near-exact solution for c. In the former case, however, where only the broad general structure of the space is of interest, there is apt to be little or no loss in choosing c larger than the best value. (Note how small the distortion is in the configuration in Figure 3 for $c = 4$.)

A very convenient, crude approximation to c in this case is as follows: Make all of the relative interpoint distances positive by adding some arbitrary positive number to every element of S_{ik}. A gross approximation to the solution would be to ignore the additive constant problem from this point on,

i.e., consider these relative interpoint distances as absolute interpoint distances and proceed with the analysis. Since the underestimation of c could lead to an imaginary space or to the addition of extraneous dimensions, this arbitrary positive number should be chosen large. If the smallest s_{jk} is felt to represent a distance judgment fairly close to a true psychological zero, the average of the s_{jk} will be a good choice for c. If the stimuli are thought to be psychologically disparate from each other, the largest s_{jk} can be chosen for c. However, just selecting a large value for the additive constant without applying the general solution given above can also be dangerous and should be done only with extreme caution. It is apparent in Figure 2 that for large values of c, some of the small roots take on appreciable values and may thus be included in the selection of the large roots. Extreme care should be taken, then, in interpreting any dimensions obtained without applying the general solution for the additive constant, for some of them may have been added in this extraneous fashion.

Procedure for Obtaining the Additive Constant in Multidimensional Scaling

1. Using any of the suggested scaling procedures, obtain a set of relative interpoint distances. Adjust these distances so that the smallest scale value is equal to zero to produce S_{jk}. Also $S_{jj} = 0$. Then construct a matrix S^2 of squared relative distances, where each element is s_{jk}^2.

2. Compute $\sum_k s_{jk}$, $\sum_j s_{jk}$, $\sum_j \sum_k s_{jk}$, $\sum_k s_{jk}^2$, $\sum_j s_{jk}^2$, $\sum_j \sum_k s_{jk}^2$,

$$\frac{1}{n}\sum_k s_{jk}, \frac{1}{n}\sum_j s_{jk}, \frac{1}{n^2}\sum_j \sum_k s_{jk}, \frac{1}{n}\sum_k s_{jk}^2, \frac{1}{n}\sum_j s_{jk}^2, \frac{1}{n^2}\sum_j \sum_k s_{jk}^2.$$

3. Substitute the proper coefficients in (17) to obtain the sum of the diagonals of B.

4. Construct matrix A, see (22), and matrix E, see (23).

5. Select a trial value for c. Usually the largest entry in S_{jk} will suffice.

6. Obtain a B matrix by inserting the trial c into (13) or (21).

7. Extract the large principal components (X_i vectors) of B. The centroid method may be used as an approximation. Iteration of the solution will then give closer approximations to principal axes as well as closer approximations to c. If, however, the trial value of c happens to be nearly the correct value, these X_i vectors will be very close to proportionality with the columns of F, the eventual matrix of projections.

8. Normalize each X_i vector so that the sum of the squares of its elements is equal to unity.

9. Compute the coefficients for the equation for the sum of the latent roots of B; see (29).

10. Set the expression for the sum of the latent roots (29) equal to the expression for the sum of the diagonals (17) found in step 3. Solve the resulting quadratic for c, selecting the value of c which produces the largest $\sum \beta$.

11. Insert this value of c into (13) or (21) to obtain a B matrix. Factor this B matrix to obtain the final F. $B = FF'$. Alternately, utilize (AX) and (EX) from step 9 and construct (BX) using (34). Then compute the matrix $(X'BX)$. The latter is symmetric and should be factored by the diagonal method to obtain K'. Compute K^{-1} and insert it into (32).

12. Iterate this solution for c by substituting columns of F for the X_i vectors in (29) above. Set this new equation for $\sum \beta$ equal to (17) and solve for a closer approximation to the additive constant. Further experience with this procedure may demonstrate that steps 1–11 give such a close approximation to c that this iteration is unnecessary.

REFERENCES

1. Abelson, R. P. A technique and a model for multidimensional attitude scaling. (In preparation.)
2. Attneave, F. Dimensions of similarity. *Amer. J. Psychol.*, 1950, **63**, 516–556.
3. Guttman, L. General theory and methods for matric factoring. *Psychometrika*, 1944, **9**, 1–16.
4. Klingberg, F. L. Studies in measurement of the relations among sovereign states. *Psychometrika*, 1941, **6**, 335–352.
5. Messick, S. J. The perception of attitude relationships: A multidimensional scaling approach to the structuring of social attitudes. Ph.D. Thesis (in preparation), Princeton University, 1954.
6. Thurstone, L. L. Multiple-factor analysis. Chicago: Univ. Chicago Press, 1947.
7. Torgerson, W. S. Multidimensional scaling: I. Theory and method. *Psychometrika*, 1952, **17**, 401–419.
8. Torgerson, W. S. A theoretical and empirical investigation of multidimensional scaling. Ph.D. Thesis, Princeton University, 1951.

Manuscript received 6/23/54

Revised manuscript received 12/3/54

PSYCHOMETRIKA—VOL. 21, NO. 1
MARCH, 1956

A RAPID NON-PARAMETRIC ESTIMATE OF MULTI-JUDGE RELIABILITY*

DESMOND S. CARTWRIGHT
UNIVERSITY OF CHICAGO

A technique is presented for obtaining a rapid estimate of reliability between judges, with special reference to qualitative judgments. It is shown that reliability and discrimination are independent and that estimates of both are needed. A method of obtaining an independent estimate of multi-judge discrimination is developed. It is shown that the size of item-samples is specified by the latter method. Tests of significance for both reliability and discrimination are described.

Whenever judgments or ratings are made, experimenters are faced with the problem of obtaining the reliability of such judgments. Frequently the average intercorrelation among the judges is computed, despite the violations of basic assumptions involved. If the judgments are not made on a metric system, however, this procedure cannot be used anyway. Three alternative methods have been proposed by Kogan and Hunt (6), employing the t statistic and analysis of variance. These authors remain unsatisfied with their methods because they still involve violations of basic assumptions. The procedures devised by Guetzkow (5) for estimating reliability in the coding of verbal material have been criticized by Tukey, who argues that Guetzkow incorrectly assumes that a "unit is either correctly classified with 100 per cent certainty, or it is classified at random" (8, p. 68). A further objection to Guetzkow's procedure is that it assumes theoretical accuracy of classification, which in many cases may be unwarranted. Whether or not a particular item is correctly classified may be a meaningful or verifiable question only in terms of some further classification or judgment, perhaps that of an expert, but still a judgment. Moreover the methods are very time-consuming, especially for three or more judges, and require sample sizes of 100-150 items. This factor may become of considerable importance in cases where several iterations of the reliability estimate are required, as in the initial building of a category system, or in the training of judges on an established system.

In the course of establishing a category system for analyzing the results

*This technique was developed in connection with research at the Counseling Center, University of Chicago. This investigation was supported by a research grant (PHS M 903) from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service. Acknowledgment is made to Dr. Lyle V. Jones, University of Chicago, for valuable discussions relevant to the technique. Responsibility rests, however, entirely with the writer.

of a role-playing experiment, the writer had need of a rapid estimate of multi-judge reliability. Because of the difficulties involved in extant procedures, and further because the units of verbal material were to be classified on a non-metric array of categories, a rapid non-parametric technique was devised.

Introduction to the Method

The general situation concerned is that where a number of persons or objects are each to be classified or rated by a number of judges in terms of some system. Such a system may be a set of diagnostic classes, role descriptions, or behavior ratings for persons. It may be any *category set*, as defined by Guetzkow, for verbal material; that is, "a number of classes or 'pigeon-holes' into which the units of qualitative data may be placed" (**5**, p. 48). Where ratings are used, the scale employed may be treated as a category set if, and only if, there is no assumption regarding the distribution of ratings. Thus a seven-point scale system may be said to have seven independent classes, named 1, 2, $\cdots$, and 7, provided all classes are equally likely to be used.

The following assumptions underlie the use of the technique to be described: (a) the items to be judged are independent; (b) the category set, or classification system, is finite; (c) the classes in the system are independent; (d) the classes are equiprobable; (e) the judges operate independently.

The basic unit is an agreement between two judges. Between J judges there are $J(J - 1)/2$ possible relations, all or some of which may be either agreements or disagreements. There are 10 possible relations between five judges. If they all agree, then the maximum number of agreements is achieved, namely, 10. If four agree together and one disagrees, then there are only $4(4 - 1)/2 = 6$ agreements. If three agree together and two agree together but differ from the first three, then there are $3(3 - 1)/2 + 2(2 - 1)/2 = 4$ agreements. The number of agreements actually achieved by J judges is to be compared with the maximum possible number of agreements. The comparison leads to a coefficient indicating the proportion of agreement achieved. This coefficient will be called alpha.

Method of Computing Alpha

Given a category set containing K classes, judges numbering J, and a sample of N items for testing the multi-judge reliability, the set of independent judgments for each item is evaluated in terms of how many judges agree and disagree. The particular combination of agreements and disagreements is called a *case*. The number of possible cases varies with J and K. For any particular J, the possible cases are constant over all values of $K \geq J$. For $K < J$, however, not all cases are possible. For example, if $J = 5$ and $K = 4$,

at least two judges must agree, so that the case of all disagree is impossible.

Since the judgments are made independently, it does not matter who in fact judges first, or who agrees with whom. The concern is solely with the number of agreements. Each of the cases for a particular J and K is associated with a specific number of agreements. In Table 1, the possible cases for $J = 5$, $K \geq 5$, are described in words and in a convenient shorthand. Also shown are the numbers of agreements associated with the various cases.

Table 1 deals with the cases for judgments on one item only. Over a sample of N items, alpha is computed as follows: (a) for every item the number of agreements achieved is tabulated; (b) these entries are then

TABLE 1

The Cases and the Number of Agreements Associated with them for $J=5$, $K \geq 5$

Cases			Agreements
#	Verbal description	Shorthand	
1	All five agree.	5:0	10
2	Four agree; one disagrees.	4:1	6
3	Three agree; two disagree with them, but agree together.	3:2	4
4	Three agree; two disagree with them, and also between themselves.	3:1:1	3
5	Two agree; two disagree with them, but agree together; one disagrees with all four others.	2:2:1	2
6	Two agree; three disagree with them, and between themselves.	2:1:1:1	1
7	All five disagree.	1:1:1:1:1	0

summed, giving the total number of agreements between the judges over N items; (c) this total is then entered as numerator over a denominator which is the maximum possible number of agreements between J judges on K classes over N items. The resultant ratio is alpha. A convenient tabular procedure is exemplified in Table 2, for which the data were taken from an actual experiment using the Thematic Apperception Test. Judges were asked to classify stories for the source of activity in the relationship between the hero and the other character. There were six possible categories: completely from self (CS); primarily from self (PS); mutuality (M); primarily from other (PO); completely from other (CO); and, no evidence (NE). The experimenter felt that the assumptions for using alpha were fulfilled.

Establishing the Probability Value of Alpha

In Table 2, alpha is seen to be .494. In practice the p-value of alpha may be found by entering Table 3 under the appropriate J and K, with alpha rounded to two decimal places. In this instance for $J = 5$, $K = 6$, alpha is found to be significant at better than the .05 level.

Use of Table 3 presupposes that the assumptions for computing alpha are fulfilled. It will be seen that independence of several judgments by any one judge is not assumed. On the contrary, it is assumed that several judgments by one judge are not independent. Such an assumption is based on the general principles of behavior. At the extremes, phenomena like response fixation warrant this assumption. In general, something like a "learned-judgment-disposition" must be posited. The concern here, therefore, is with the mean judging performance of judges. More specifically, the alpha computed over a sample of items represents the mean case of agreement achieved by the judges as a group.

The underlying theory of case-probability and the manner in which Table 3 was prepared will be briefly described.

The relative frequency of the various cases for any J and K may be expressed as

$$\frac{\textit{number of different ways of obtaining a particular case}}{\textit{number of different ways of obtaining all possible cases}}.$$

Since it is immaterial which category the first judge selects, the algebraic denominator of the foregoing expression becomes

$$K^{J-1}. \tag{1}$$

The several cases for any J and K may be equivalently expressed as the presence or absence of partitioning among the judges. The case of perfect agreement is equivalently expressed as the absence of partitioning. The case 4:1 for $J = 5$ may be expressed as the partitioning of J into two groups, say (a, b, c, d) and (e). The assumption that the K classes are equiprobable

TABLE 2

Example of Computing Alpha on Data Taken From a Thematic Apperception Test Experiment with J=5, K=6, N=18

Items	Cases						
	5:0	4:1	3:2	3:1:1	2:2:1	2:1:1:1	1:1:1:1:1
1						/	
2			/				
3				/			
4		/					
5	/						
6					/		
7						/	
8	/						
9		/					
10	/						
11						/	
12		/					
13		/					
14				/			
15	/						
16				/			
17						/	
18		/					
Sum =	4	5	1	3	1	4	0
Agreements per case =	10	6	4	3	2	1	0
Product =	40	30	4	9	2	4	0

Sum products = 89. Maximum agreements = 18 x 10 = 180

Alpha = 89/180 = .494.

is equivalent to the assumption that the judges are equally likely to select any of the K classes. Under these circumstances, the probability that J judges will select exactly r of the possible K classes can be shown to be (2, p. 26)

$$\frac{1}{K^J} \cdot \frac{K!}{(K-r)!} \cdot \frac{\Delta^r 0^J}{r!}. \tag{2}$$

9-374
7.9.70

TABLE 3

Values of Alpha Required for Various Levels of Probability when J = 3, 4, and 5, and K = 4 - 16

	J=3				J=4					J=5						
K \ P	.10	.05	.025	.01	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005	.002	.001
4	.95	-	-	-	.76	.90	.98	-	-	.46	.64	.85	.96	1.00	-	-
5	.91	.99	-	-	.62	.83	.93	1.00	-	.38	.49	.70	.89	.96	1.00	-
6	.88	.97	-	-	.32	.75	.89	.97	1.00	.35	.40	.55	.81	.91	.98	1.00
7	.85	.95	1.00	-	.32	.65	.84	.95	.99	.33	.37	.40	.69	.85	.95	.99
8	.82	.93	.99	-	.31	.55	.79	.93	.98	.31	.37	.40	.57	.77	.92	.97
9	.80	.92	.98	-	.31	.33	.73	.90	.96	.27	.35	.39	.48	.68	.88	.95
10	.77	.90	.97	1.00	.31	.32	.66	.88	.94	.23	.33	.38	.40	.58	.83	.92
11	.75	.89	.96	1.00	.30	.32	.59	.85	.93	.20	.31	.37	.40	.51	.77	.89
12	.73	.88	.95	1.00	.30	.32	.51	.82	.92	.20	.31	.36	.40	.43	.70	.86
13	.70	.86	.94	.99	.30	.32	.33	.78	.90	.20	.29	.36	.39	.40	.62	.82
14	.68	.85	.93	.99	.29	.32	.33	.75	.88	.19	.26	.35	.39	.40	.57	.77
15	.66	.84	.93	.98	.29	.32	.33	.71	.86	.19	.23	.34	.38	.40	.52	.73
16	.63	.83	.92	.98	.28	.31	.33	.67	.84	.19	.20	.33	.38	.40	.45	.65

For the case of perfect agreement $r = 1$; all judges select the same one class. Hence the third term of (2) becomes unity, and the remaining terms reduce to

$$\frac{1}{K^J} \cdot \frac{K!}{(K-r)!} \cdot 1 = \frac{K}{K^J} = \frac{1}{K^{J-1}}. \tag{3}$$

The probability of the case of perfect agreement is seen to be the

reciprocal of the total number of different ways in which J judges may behave to obtain all possible cases, namely, the reciprocal of (1).

Cases involving disagreement are those in which judges may be said to be partitioned into two or more groups. As an example, the case 2:1:1:1 for $J = 5$ may be expressed as the partitioning of five judges into four groups. The probability of this case for $J = 5$ and $K = 5$ is given by substituting in (2) with $r = 4$, yielding

$$\frac{1}{5^5} \cdot \frac{5!}{(5-4)!} \cdot \frac{\Delta^4 0^5}{4!}.$$

The right-hand term of (2) has been tabulated by Fisher (**2**, p. 78), so that computation of leading rth differences of the Jth powers of natural numbers commencing at zero is unnecessary. Using Fisher's table for the present case 2:1:1:1 for $J = 5$ and $K = 5$ the third term is found to be 10, and the probability for the case is found to be .38400.

There are instances in which two or more different cases may each be represented as the partitioning of J judges into r groups. For $J = 5$ and $K = 5$ the cases 4:1 and 3:2 each partition J into two groups but differ as to size of the groups. Since (2) gives the probability for the total partition, there is need to separate the respective probabilities for each case. This is done by the usual method of proportional parts.

Following Whitworth (**9**, pp. 58-9), the number of ways in which $J = (x + y)$ judges may be divided into two groups so that one contains x and the other y judges is

$$(x + y)!/(x!y!). \tag{4}$$

If J is an even number, the situation may arise where $x = y$. For $J = 4$ the case 2:2 would be one instance of this. There are six possible partitions of the form (x) (y):

$$(ab)(cd); \quad (ac)(bd); \quad (ad)(bc);$$
$$(cd)(ab); \quad (bd)(ac); \quad (bc)(ad).$$

Except for the assignment to x or y, the second three partitions duplicate the first three. Since the x or y assignment is irrelevant for the present purposes, when $x = y$ a correction factor must be introduced to divide out duplicate partitions. This may be achieved by taking the factorial of the number of numerically equivalent groups in the partition. Let that number be e. Then, for $J = 4$ a partition of 2:2 has $e = 2$. For $J = 5$ a partition of 2:1:1:1 has $e = 3$. The general expression required for dividing $J = (x + y)$ judges into two groups is

$$J!/(x!y!e!). \tag{5}$$

Where $e = 0$, $e! = 1$, in which case (4) is unchanged.

Since the right-hand member of (2) gives the number of ways of dividing J judges into r groups of all possible sizes, and since (5) gives the number of ways of dividing J judges into two groups of any particular size, the summation of (5) over all possible group sizes for $J = x + y$ and $r = 2$ will yield the identity

$$\sum_{i=1}^{n} J!/(x_i!y_i!e_i!) = \Delta^2 0^J/2!. \tag{6}$$

For $J = 5$ and $K = 5$ the cases 4:1 and 3:2 both entail the partitioning of J into $r = 2$ groups. In both cases $e = 0$. Substituting in the left-hand side of (6) yields $5 + 10$. Entering Fisher's table for $\Delta^r 0^J/r!$ with $r = 2$ and $J = 5$, the result is 15. From (2) the total probability associated with the two cases is .09600. The probabilities associated with each case are therefore

$$(4{:}1): \qquad .09600 \times 5/15 = .03200$$

and

$$(3{:}2): \qquad .09600 \times 10/15 = .06400.$$

Where J judges are to be partitioned into $r = 3$ groups, (6) becomes

$$\sum_{i=1}^{n} J!/(x_i!y_i!z_i!e_i!) = \Delta^3 0^J/3!. \tag{7}$$

For $J = 5$ and $K = 5$ the cases 3:1:1 and 2:2:1 both entail the partitioning of J into $r = 3$ groups. In both cases $e = 2$. The probability for the total partition is given by (2) as .48000. Substituting in (7) yields $10 + 15 = 25$. Hence the probabilities associated with each case are

$$(3{:}1{:}1): \qquad .48000 \times 10/25 = .19200$$

and

$$(2{:}2{:}1): \qquad .48000 \times 15/25 = .28800.$$

From (2), in conjunction with (6) and (7), or the extensions of the latter for any r, the exact probabilities of all cases for any J and K may be derived. The set of probabilities for $J = 5$ and $K = 5$ are shown in Table 4.

As may be seen from Table 4, the p-values are associated with particular cases. The alpha coefficients for these cases are exact for a set of judgments on one item. In practice, however, an alpha coefficient represents the mean case of agreement achieved over N items. Under these conditions, alpha may vary continuously between 0.00 and 1.00. The probabilities associated with values of alpha intermediate between two adjacent cases may be assumed to lie correspondingly between the two associated probability values. Thus, for $J = 5$ and $K = 5$ an alpha of .57 would lie between cases 2 and 3, namely,

between .60 and .40 (cf. Table 4). The p-value of that alpha would lie between the p-values associated with cases 2 and 3, namely, between .03200 and .06400. An estimate of the p-value for alpha may be made by linear interpolation. For alpha = .57, in the present example, $p = .03200 + 3(.06400 - .03200)/20 = .03680$.

Table 3, giving the values of alpha required for various levels of prob-

TABLE 4

The Cases, Alpha Coefficients, and Probability Values for J=5, K=5

#	Shorthand	Agreements	Alpha	P-values
1	5:0	10	1.00	.00160
2	4:1	6	.60	.03200
3	3:2	4	.40	.06400
4	3:1:1	3	.30	.19200
5	2:2:1	2	.20	.28800
6	2:1:1:1	1	.10	.38400
7	1:1:1:1:1	0	.00	.03840

ability for $J = 3$ to 5 and $K = 4$ to 16, was prepared on the basis of the argument in the foregoing paragraphs. For a particular value of J, the probabilities associated with each possible case over the values of K from 4 to 16 were computed directly and set out in tabular form. The column of p-values in Table 4 is typical of this preliminary tabulation. In the preliminary tables, thirteen such columns appeared, one for each value of K. Values of alpha intermediate between adjacent cases were inserted into the preliminary table, with successive intervals of .01. Probabilities for the intermediate values of alpha were computed by linear interpolation between the p-values of adjacent cases. The interpolations were taken to five decimal places. From the preliminary tables, the values of alpha having p-values at or slightly less than .10, .05, .025, etc., were abstracted to form Table 3.

Sample Size for Items

The question of sample size is important in two ways. First, the nature of much work in psychology precludes the possibility of obtaining large samples, particularly for pilot reliability studies. Most experimenters cannot easily obtain 150 Rorschach protocols on which to try out a new category

system. Second, the size of samples is important for the methodological problems associated with judges' discrimination among the classes of the system. It will be shown below that the minimum sample size required is equal to three times the number of classes in the system, that is, N (min.) = $3K$. This minimum is required by the method of obtaining an estimate of judges' discrimination. It also serves, however, as a useful lower limit of sample size for estimating alpha.

Sample Size and Judge Discrimination

There are four possible extreme situations in the relations between multi-judge agreement and multi-judge discrimination. These are: maximum agreement and maximum discrimination; maximum agreement and minimum discrimination; minimum agreement and maximum discrimination; minimum agreement and minimum discrimination. The first of these situations is exemplified, for $J = 3$, $K = 4$, and $N = 12$, by the judges achieving perfect agreement on all items with three items placed in $K1$, three in $K2$, three in $K3$, and three in $K4$. This situation gives rise to a theorem which states that for a reasonable estimate of discrimination the sample must number a multiple of K. To prove this theorem note that the case of perfect agreement can be equipossible among the K classes if, and only if, the sum of possible cases (that is, the number of items available for classification in any K) is the same for all K. In the technique described below, it will be seen that, for maximum discrimination, chi-square is always zero. This is possible under all conditions only if the number of items is equal to a multiple of K.

The situation of maximum agreement and minimum discrimination would be exemplified for $J = 3$, $K = 4$, and $N = 12$, by the judges agreeing perfectly on all items, but with all items placed only in one class, say $K1$. In other words, though the judges agree perfectly, it may be due to the fact that all the items look the same to them—they all look like $K1$.

In the attempt to develop a measure of multi-judge discrimination, the methods of measuring the amount of transmitted information proposed by Garner and Hake (4) were examined. It was not found possible to use any of these methods. But the comparability between these methods and those of analysis of variance, which is pointed out by the authors, suggested the possibility of using rank analysis of variance. The technique to be described adapts analysis of variance by ranks, as developed by Pitman (7), Friedman (3), and Wilcoxon (9) to the kind of data with which the present problem of discrimination is concerned. In this technique good multi-judge discrimination is indicated by a non-significant value of chi-square. Poor discrimination is indicated by a significant value of chi-square. For the situation of minimum agreement and minimum discrimination, where each judge places all items in one class and no two judges use the same class, it is possible for a non-

significant chi-square to result from certain J and K values unless the number of items (or replications) is greater than $2K$. Since, from the argument above, the number of items in the sample must be a multiple of K, the minimum number of items in a sample must therefore be equal to $3K$.

Estimating the Multi-Judge Discrimination

The method of estimating the multi-judge discrimination is the method of analysis of variance by ranks. A convenient exposition of the method and formula for computing chi-square, together with a reproduction of the chi-square chart developed by Bliss (1), is presented by Wilcoxon (9). The method of preparing the data for this analysis in regard to judge discrimination is as follows:

A table is prepared having $K \times N$ cells. The classes of the K system are considered as treatments. The items of the sample are considered as replications. On each item the number of judges placing the item in a particular category or class is entered in the appropriate cell for that item row and class column. The entries are made for all items. Each row is then ranked, with the highest rank being assigned to the cell (class) having the largest number of judges, and the lowest rank being assigned to the cell having the least number of judges. Ties are scored in the usual way. The columns are then summed, the sums are squared, and the squares are summed. This last figure is entered in the equation for chi-square.

An example of the procedure is given in Table 5. The data for this example are the same as those for Table 2.

From Table 5 it is seen that the chi-square has a p-value of .183. This value is clearly non-significant and adequately represents the distribution of class selections by the judges. There is little or no concentration on one or a few classes, and hence the discrimination is good. In other instances, some concentration may appear and be manifested in a value of chi-square approaching significance. In such an instance, where for example the p-value was .075, it would be a matter for the experimenter to decide whether or not the discrimination is good enough, in terms of the kind of items presented and other factors.

In general, p-values between .10 and .06 will require careful consideration in relation to the alpha achieved by the judges for agreement. Values of chi-square which have a p-value greater than .10 may be taken as indicating good discrimination on the part of the judges. Values of chi-square which have a p-value of .05 or less may be taken to represent poor discrimination. In such circumstances it may be possible to combine two or more classes, thus raising discrimination efficiency on a reduced number of classes. Such a course would require recomputation of alpha on the revised K system.

Since alpha varies independently of judge discrimination, a high value of alpha should always be checked for its meaningfulness in terms of whether

TABLE 5

Example of Estimating Multi-Judge Discrimination on Data Taken from a Thematic Apperception Test Experiment with J=5, K=6, N=18

Item	Classes											
	CS Rank		PS Rank		M Rank		PO Rank		CO Rank		NE Rank	
1	0	1+	1	4	1	4	2	6	0	1+	1	4
2	3	6	2	5	0	2+	0	2+	0	2+	0	2+
3	0	2	0	2	3	6	1	4+	1	4+	0	2
4	0	2+	4	6	0	2+	0	2+	1	5	0	2+
5	0	3	0	3	0	3	0	3	0	3	5	6
6	1	4	2	5+	0	2	0	2	0	2	2	5+
7	1	4	2	6	1	4	1	4	0	1+	0	1+
8	0	3	0	3	0	3	5	6	0	3	0	3
9	1	5	4	6	0	2+	0	2+	0	2+	0	2+
10	0	2+	0	2+	0	2+	1	5	4	6	0	2+
11	1	4	1	4	1	4	2	6	0	1+	0	1+
12	1	5	4	6	0	2+	0	2+	0	2+	0	2+
13	4	6	1	5	0	2+	0	2+	0	2+	0	2+
14	0	2	1	4+	1	4+	3	6	0	2	0	2
15	0	3	0	3	0	3	0	3	0	3	5	6
16	0	2	0	2	3	6	1	4+	1	4+	0	2
17	1	4	1	4	1	4	2	6	0	1+	0	1+
18	0	2+	4	6	0	2+	0	2+	1	5	0	2+
Totals	62.0		77.5		61.0		71.0		54.0		52.5	
Squares	3844.0		6006.25		3721.0		5041.0		2916.0		2756.25	

Sum of squares = 24284.50

$$X^2_r = \frac{12}{np(p+1)} \times \text{sum (rank totals)}^2 - 3n(p+1)$$

$$= \frac{12}{18 \times 6 \times 7} \times 24284.50 - 3 \times 18 \times 7 = 7.468$$

Degrees of freedom = p-1 = 5. Probability of $X^2_r = 7.468 = .183$.

or not the judges showed good discrimination also. If they did not, it is possible, though not necessary, that their degree of agreement was related to their poor discrimination.

REFERENCES

1. Bliss, C. I. A chart of the chi-square distribution. *J. Amer. statist. Assoc.*, 1944, **39**, 246–248.
2. Fisher, R. A., and Yates, F. Statistical tables for biological, agricultural and medical research. (4th Ed.), New York: Hafner, 1953.
3. Friedman, M. The use of ranks to avoid the assumption of normality. *J. Amer. statist. Assoc.*, 1937, **32**, 675–701.
4. Garner, W. R., and Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.
5. Guetzkow, H. Unitizing and categorizing problems in coding qualitative data. *J. clin. Psychol.*, 1950, **6**, 47–58.
6. Kogan, L. S., and Hunt, J. McV. Problems of multi-judge reliability. *J. clin. Psychol.*, 1950, **6**, 16–19.
7. Pitman, E. J. G. Significance tests which may be applied to samples from any populations. III. The analysis of variance test. *Biometrika*, 1937, **29**, 322–335.
8. Tukey, J. W. Discussion of symposium. *J. clin. Psychol.*, 1950, **6**, 61–74.
9. Whitworth, W. A. Choice and chance. New York: Stechert, 1934.
10. Wilcoxon, F. Some rapid approximate statistical procedures. New York: American Cyanamid Co., 1949.

Manuscript received 11/22/54
Revised manuscript received 2/25/55

PSYCHOMETRIKA—VOL. 21, NO. 1
MARCH, 1956

A STUDY OF SPEED FACTORS IN TESTS AND ACADEMIC GRADES*

FREDERIC M. LORD

EDUCATIONAL TESTING SERVICE

Speeded and unspeeded tests of vocabulary, spatial relations, and arithmetic reasoning were factorially analyzed, together with certain reference tests and academic grades. Lawley's maximum likelihood method was used, the computations being carried out on the Whirlwind electronic computer. Four different speed factors were isolated, together with a second-order general speed factor. Consistent small positive correlations between the academic grades and the speed factors were found.

The speed with which an examinee responds to the items in a test frequently affects his score. Almost all achievement and aptitude tests are to some extent measures of "speed." Tests for factor analyses are frequently speeded because many tests must be given in a limited time.

Much remains to be learned about "speed," in spite of the fact that it is commonly an element in test scores. Is speed on cognitive tests a unitary trait? Or are there different kinds of speed for different kinds of tasks? If so, how highly correlated are these different kinds of speed? How highly correlated are speed and level on the same task? How do various criteria relate to speed, and how speeded should tests to predict these criteria be? These are the questions which the present study attacks.

Some Previous Results

Factor analytic studies have often isolated a "perceptual-speed factor," usually measured by tests requiring simple, rapid visual discriminations. "This factor is characterized by the task of (quickly) finding in a mass of distracting material a given configuration which is borne in mind during the search" (6). Any speed test composed of very easy items is likely to have a loading on this factor. A more recent publication (7) breaks down "perceptual speed" into at least two factors, "speed of symbol discrimination" and "form

*The writer is indebted to Dr. John French, to Dr. David Saunders, and especially to Dr. Ledyard R Tucker for helpful suggestions and theoretical advice throughout the course of this study. The active cooperation of Dr. William Shields, Educational Advisor, and of many others at the United States Naval Academy at Annapolis has been invaluable. The author is very grateful to Dr. P. Youtz and Dr. C. W. Adams for the opportunity to use Whirlwind I, a high-speed computer sponsored by the Office of Naval Research, and to Dr. H. Denman for help is programming and in putting the program on the computer. He also wishes to express his deep appreciation to Dr. Hubert Brogden and Miss Bertha Harper of The Adjutant General's Office for the opportunity to use their matrix rotator and for helpful guidance in its operation.

perception," the former relating to familiar symbols, the latter to unfamiliar figures.

Other factors related to speed include finger dexterity, fluency of expression, ideational fluency, reaction time, speed of association, speed of judgment, tapping, word fluency (6). Speed of closure and motor speed are included in (7). Rimoldi (20) finds a "speed of judgment," a "speed of cognition," and a second-order "personal tempo" factor; but his subjects, like those in many earlier studies, were to work at a "natural, congenial" speed rather than at the maximal speed required by most tests.

Since many tests in factor analytic studies are speeded, many of the factors are speed factors, although not always so described. An example is the "number" factor, which is commonly measured by highly speeded tests of addition, subtraction, multiplication, and division. This factor will here be referred to as the *number-speed factor*.

In spite of the presence of both speeded and unspeeded tests in most factor analysis batteries, a general intellectual-speed factor has not routinely been found. Studies designed to investigate the existence of both general and specific speed factors in ordinary aptitude test batteries have been few and have yielded conflicting evidence (3, 4, 17, 18, 21, 22).

For further consideration of "speed factors," the reader is referred to (24, pp. 80-85) and to the 33 references in (8).

Data for the Present Study

The Subjects

All measures in this study were obtained on 649 students in the entering class at the United States Naval Academy at Annapolis. This large number of cases was used to obtain clearly interpretable results.

The Tests

The study centers around tests of the verbal factor, of spatial ability, and of arithmetic reasoning, because of the widespread use of tests in these areas.

In each area, seven tests were administered. One was the regular admissions examination, denoted by (A), which is only slightly speeded. The remaining six were short experimental tests administered at the beginning of the school year. These were parallel in content, but different in degree of speededness. Two were "level" tests, denoted by (L), involving virtually no speed. One was moderately speeded (M). The remaining three tests were highly speeded (S). In order to confound practice effect insofar as possible, the tests were administered in scrambled order, as follows: LSMSLS. The examinee was told the degree of speededness that would be required.

Six reference-factor tests (number, perceptual speed, word fluency) also were administered. These are designated by (R).

A more complete description of all the tests follows.

1. *Word Fluency* (*R*). The examinee writes as many words and their opposites as he can in four minutes. This test was included so as to determine its relation to the verbal factor and to the verbal-speed factor, if such were found.

2. *Verbal* (*A*). This test contained both word-analogies and "double-definitions" items. The latter item type is essentially a sentence with two missing words to be selected from alternative pairs of words provided, thus producing a simple definition of one of the missing words.

3, 4. *Vocabulary* (*L*). These tests require finding among the choices a word opposite in meaning to the given key word. Also 5. *Vocabulary* (*M*) and 6, 7, 8. *Vocabulary* (*S*). [9. *Vocabulary* (*LIA*) is merely the "last-item-attempted score" on test 7, to be discussed below.]

10. *Spatial Relations* (*A*) contained block-counting and "identical-blocks" items. The latter require the examinee to indicate which of five drawings represents a key block drawn from a different angle.

11, 12. *Intersections* (*L*). These tests require the examinee to visualize the two-dimensional outline of the intersection of a solid geometric object cut by a plane. Also 13. *Intersections* (*M*) and 14, 15, 16. *Intersections* (*S*). [17. *Intersections* (*LIA*) is merely the "last-item-attempted score" on test 15.]

18. *Mathematics* (*A*) is composed of arithmetic reasoning, algebra, and geometry items.

19, 20. *Arithmetic Reasoning* (*L*) consist entirely of the usual arithmetic-reasoning items. Also 21. *Arithmetic Reasoning* (*M*) and 22, 23, 24. *Arithmetic Reasoning* (*S*). [25. *Arithmetic Reasoning* (*LIA*) is merely the "last-item-attempted score" on test 23.]

26, 27. *Number Speed* (*R*) are highly speeded reference tests for the number-speed factor. 26 consists of simple addition and division, 27 of easy subtraction and multiplication.

28, 29, 30. *Perceptual Speed* (*R*) are reference tests for the perceptual-speed factor. 28. *Cancellation* requires the examinee to cross out as many letter *A*'s in a paragraph as he can in two minutes. 29. *Picture Discrimination* requires him to indicate which of three very sketchily drawn faces is different from the other two. 30. *Number Checking* requires him to indicate whether two multi-digit numbers are the same or different.

TABLE 1

Background Information and Data on Speededness for the "Experimental" Tests

	Tests	Speed-edness	Number of Items	Test-ing Time	Items per Hour	Per cent* of Examinees Finishing
3,4	Vocabulary	L	15	7	129	97
5	"	M	30	5	360	71
6,7,8	"	S	75	5	900	2
11,12	Intersections	L	15	20	45	98
13	"	M	20	12	100	75
14,15,16	"	S	35	9	233	11
19,20	Arithmetic Reasoning	L	10	20	30	94
21	" "	M	15	15	60	50
22,23,24	" "	S	30	10	180	4

*The mean of two values in the case of the level tests, of three in the case of the speed tests.

Table 1 summarizes the background information about the "experimental" tests and shows the proportion of examinees who answered the last item in each test. The speeded tests were in fact very highly speeded. There is reason to believe that many or all of the examinees who answered the last item of the speeded tests skipped many items or responded at random.

Scoring

The three admissions tests are composed of multiple-choice items having five (in a few cases, eight) alternative responses. The score obtained for each test was the number of items answered correctly.

The eighteen experimental vocabulary, intersections, and arithmetic-reasoning tests were all composed of five-choice items and were scored number-right minus one-fourth-number-wrong. This "correction for guessing" was made in order that any speed factor that might be found should not be open to the challenge that it was merely a "willingness-to-guess-wildly" factor. It would have been wrong to include both corrected and uncorrected scores on the same test in a straightforward factor analysis, because of their experimental dependence. Some further investigation of the effect of the correction for guessing was nevertheless planned. For this purpose, number-right (NR) scores were obtained for tests 7, 15, and 23, these new scores being designated as variables 37, 38, and 39.

The score on each of the six reference tests was the number of right answers, because this is the usual method for scoring these tests. Scored in any other way, they might no longer represent the same reference factors.

In addition to the regular score, a "last-item-attempted score" (LIA) for one speeded test in each of the three areas gives a crude measure of rate-of-work. Inclusion of such scores in the present study was considered desirable, although in general the study is primarily concerned with the type of scores normally used in work with aptitude tests. The statistical method used to deal with the experimental dependence of these scores and of the NR scores on the other scores obtained from the same tests will be outlined later.

School Grades

During their first year, all students at Annapolis normally receive grades in each of the following:

31. *English Composition and Literature.*

32. *Foreign Language.* (Each student selects one of several available.)

33. *Engineering Drawing and Descriptive Geometry.*

34. *Chemistry.*

35. *Mathematics.* (Plane trigonometry, college algebra, plane and solid analytic geometry, and calculus.)

36. *Conduct.* (The method by which these grades are assigned need not concern us, since no factor loadings of interest were found for this variable.)

In the present study, each numerical grade is averaged over two semesters. Each semester course grade represents a combination of day-to-day course work and final-examination performance weighted in the ratio of three to two. The instructors could not have had knowledge of the test scores, with the possible exception of the three admissions tests.

The final examinations were virtually unspeeded, almost every student finishing. The day-to-day work in class varied but was not in general compulsorily speeded. It is not known whether students felt pressed for time while doing their homework assignments.

Statistical Analysis

Normalizing

All variables were normalized before product-moment correlations were computed. This was considered desirable since otherwise any speed factors that might be found might conceivably have been attributable to certain common features in the shapes of the score distributions of the speeded tests (e.g., skewness), rather than to a real speed factor.

The Correlations

The use of product-moment correlations is required by the significance tests to be described later. The correlation matrix is presented as Table 2.

TABLE 2
Matrix of Intercorrelations (decimal point omitted)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	---	264	200	262	249	281	341	321	314	056	111	100	064	136	080	090	122	181
2	264	---	720	720	790	666	732	679	388	178	174	135	165	153	138	126	063	413
3	200	720	---	669	706	620	693	641	328	134	150	055	102	119	093	079	087	287
4	262	720	669	---	690	648	697	650	343	126	181	144	184	171	128	121	054	310
5	249	790	706	690	---	660	745	700	393	136	111	082	115	093	082	075	046	326
6	281	666	620	648	660	---	775	757	531	176	162	137	176	188	174	161	140	323
7	341	732	693	697	745	775	---	855	(671)	135	122	061	103	127	107	100	130	337
8	321	679	641	650	700	757	855	---	609	138	094	067	117	135	104	114	126	334
9	314	388	328	343	393	531	(671)	609	---	078	060	026	031	127	086	109	291	290
10	056	178	134	126	138	176	135	138	078	---	480	541	548	543	534	553	236	349
11	111	174	150	181	111	162	122	094	060	480	---	722	714	696	701	696	270	312
12	100	135	055	144	082	137	061	067	026	541	722	---	767	714	738	750	246	255
13	064	165	102	184	115	176	103	117	031	548	714	767	---	748	760	770	262	274
14	136	153	119	171	093	188	127	135	127	543	696	714	748	---	796	788	428	266
15	080	138	093	128	082	174	107	104	086	534	701	738	760	796	---	833	(481)	253
16	090	126	079	121	075	161	100	114	109	553	696	750	770	788	833	---	449	295
17	122	063	087	054	046	140	130	126	291	236	270	246	262	428	(481)	449	---	132
18	181	413	287	310	326	323	337	334	290	349	312	255	274	266	253	295	132	---
19	156	341	300	316	265	276	291	273	192	290	337	295	337	322	312	335	105	555
20	165	324	251	311	258	290	270	288	176	271	357	302	362	300	286	314	090	596
21	154	363	285	339	299	323	313	348	208	331	295	297	344	309	289	324	121	638
22	183	335	279	255	300	357	353	384	303	309	289	280	334	310	291	361	174	628
23	174	329	275	286	297	333	350	347	282	321	337	280	338	345	338	370	192	602
24	216	364	276	301	314	380	364	388	316	275	257	230	252	266	254	307	152	596
25	233	195	160	147	194	226	294	290	455	171	124	112	105	192	177	194	394	355
26	258	051	047	013	086	133	196	205	339	027	-046	-068	-092	-018	-028	004	102	260
27	200	050	044	013	088	165	178	202	358	058	-018	-022	-045	012	-028	037	070	351
28	223	058	052	043	097	211	223	233	316	068	103	076	065	114	072	123	150	160
29	212	121	084	085	100	245	254	250	354	247	215	243	231	292	283	277	231	194
30	122	-062	-038	-023	-056	093	105	112	249	-016	-028	-049	-080	-048	-022	005	052	113
31	303	560	497	568	537	519	590	540	373	120	143	093	102	073	037	058	003	302
32	184	210	172	205	192	186	220	226	204	082	136	125	100	110	100	106	019	177
33	170	184	084	186	138	247	192	221	182	476	546	582	605	574	568	584	225	308
34	249	230	172	238	196	270	258	248	228	238	321	310	318	327	292	320	141	376
35	207	156	119	145	128	213	211	210	258	218	328	288	300	308	283	326	143	435
36	-082	-051	-106	-017	-064	-034	-038	-019	-020	055	019	025	056	023	023	037	-054	059
37	346	707	665	673	724	770	(986)	856	(745)	124	111	049	087	127	102	099	171	334
38	097	135	099	124	082	181	120	118	139	522	678	698	722	790	(977)	825	(625)	256
39	216	325	257	271	295	341	362	362	368	326	329	272	317	352	340	366	297	597

TABLE 2 (cont.)
Matrix of Intercorrelations (decimal point omitted)

0	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	
5	154	183	174	216	233	258	200	223	212	122	303	184	170	249	207	-082	346	097	216	1
4	363	335	329	364	195	051	050	058	121	-062	560	210	184	230	156	-051	707	135	325	2
1	285	279	275	276	160	047	044	052	084	-038	497	172	084	172	119	-106	665	099	257	3
1	339	255	286	301	147	013	013	043	085	-023	568	205	186	238	145	-017	673	124	271	4
8	299	300	297	314	194	086	088	097	100	-056	537	192	138	196	128	-064	724	082	295	5
0	323	357	333	380	226	133	165	211	245	093	519	186	247	270	213	-034	770	181	341	6
0	313	353	350	364	294	196	178	223	254	105	590	220	192	258	211	-038	(986)	120	362	7
8	348	384	347	388	290	205	202	233	250	112	540	226	221	248	210	-019	856	118	362	8
6	208	303	282	316	455	339	358	316	354	249	373	204	182	228	258	-020	(745)	139	368	9
1	331	309	321	275	171	027	058	068	247	-016	120	082	476	238	218	055	124	522	326	10
7	295	289	337	257	124	-046	-018	103	215	-028	143	136	546	321	328	019	111	678	329	11
2	297	280	280	230	112	-068	-022	076	243	-049	093	125	582	310	288	025	049	698	272	12
2	344	334	338	252	105	-092	-045	065	231	-080	102	100	605	318	300	056	087	722	317	13
0	309	310	345	266	192	-018	012	114	292	-048	073	110	574	327	308	023	127	790	352	14
6	289	291	338	254	177	-028	-028	072	283	-022	037	100	568	292	283	023	102	(977)	340	15
4	324	361	370	307	194	004	037	123	277	005	058	106	584	320	326	037	099	825	366	16
0	121	174	192	152	394	102	070	150	231	052	003	019	225	141	143	-054	171	(625)	297	17
6	638	628	602	596	355	260	351	160	194	113	302	177	308	376	435	059	334	256	597	18
8	606	551	532	514	211	155	193	100	115	026	309	192	355	424	411	072	277	295	507	19
-	560	544	548	534	229	207	276	088	129	039	281	228	328	429	475	029	258	270	518	20
0	---	632	578	571	284	187	310	147	155	090	318	177	313	378	376	047	304	278	554	21
4	632	---	632	639	393	320	409	178	195	120	278	190	338	380	416	083	351	293	629	22
8	578	632	---	610	(454)	241	321	178	242	148	272	225	350	420	430	095	347	332	(949)	23
4	571	639	610	---	349	329	407	175	199	137	300	213	302	364	400	084	360	252	596	24
9	284	393	(454)	349	---	322	304	245	280	187	178	157	183	280	300	-012	340	237	(670)	25
7	187	320	241	329	322	---	646	347	316	443	160	222	081	236	342	007	227	000	286	26
3	310	409	321	407	304	646	---	322	264	464	167	198	090	234	359	067	212	-013	348	27
3	147	178	178	175	245	347	322	---	370	428	152	175	207	194	259	-033	248	094	218	28
9	155	195	242	199	280	316	264	370	---	392	109	084	331	201	216	021	275	291	290	29
9	090	120	148	137	187	443	464	428	392	---	082	127	123	165	287	047	135	-009	170	30
4	318	278	272	300	178	160	167	152	109	082	---	467	300	448	386	081	584	036	277	31
3	177	190	225	213	157	222	198	175	084	127	467	---	302	497	504	147	230	094	230	32
3	313	338	350	302	183	081	090	207	331	123	300	302	---	604	570	177	194	540	351	33
9	378	380	420	364	280	236	234	194	201	165	448	497	604	---	806	125	259	283	427	34
4	376	416	430	400	300	342	359	259	216	287	386	504	570	806	---	134	221	275	438	35
4	047	083	095	084	-012	007	067	-033	021	047	081	147	177	125	134	---	-040	007	082	36
4	304	351	347	360	340	227	212	248	275	135	384	230	194	259	221	-040	---	126	381	37
3	278	293	332	252	237	000	-013	094	291	-009	036	094	540	283	275	007	126	---	359	38
9	554	629	(949)	596	(670)	286	348	218	290	170	277	230	351	427	438	082	381	359	---	39

Variables 9, 17, 25, 37, 38, and 39, which were not used in the factoring, are experimentally dependent on variables 7, 15, and 23. The consequent spuriously high correlations are placed in parentheses in the table.

Lawley's Maximum Likelihood Method of Factor Analysis

Factors were extracted by Lawley's maximum likelihood method. Since this important method has not often been mentioned in this country and since it (or a modification) is likely to become widely used in the near future, some references are listed. The basic development was given in (**12**). Extensions and further developments appeared in (**13**, **15**, **1**, **2**, and **23**). [A maximum likelihood Method II, avoiding the assumption of multivariate normality, was developed in (**13**). This second method will not be considered here, since the usual optimum properties of maximum likelihood estimates do not appear to hold (**1**, **11**). Whittle (**25**) derived a relatively simple solution for a similar situation in the special case where the variables are of known reliability.] Henrysson (**10**) reported an empirical sampling study supporting Lawley's test of significance. Bartlett (**1**, **2**) gave a significance test that is superior to Lawley's whenever the number of examinees is not large compared to the number of variables. Some recent papers (in English) by Bartlett, Lawley, and others appeared in (**26**). Very recently Rao (**19**) discussed the basic differences between Hotelling's principal-component analysis and common-factor analysis and described further developments related to those of Lawley.

Lawley's method and Thurstone's centroid method are both concerned with estimating common-factor loadings, specific-factor variance being systematically set aside. Certain characteristics of the maximum likelihood method are:

1. The number of common factors is tentatively hypothesized in advance.
2. The procedure in effect determines the population correlation matrix, having the hypothesized rank, for which the likelihood of occurrence of the observed sample in the course of random sampling is a maximum. The matrix of factor loadings exactly reproducing this matrix of population correlation coefficients is the basic result obtained by the maximum likelihood method. The result is obtained by iterative procedures.
3. The usual matrix of residuals is computed; a rigorous large-sample significance test is made to determine whether or not the residuals may plausibly be attributed solely to sampling fluctuations in the correlation coefficients.
4. If the residuals are statistically significant, the research worker repeats the foregoing process, starting with different tentative hypotheses as to the number of common factors required to explain the data, until he is ready finally to accept one of these hypotheses.
5. The usual problem of estimating the communalities ceases to be a serious cause for concern, since the maximum likelihood estimates of the communalities are one of the outcomes of the procedure.

The practical application of the maximum likelihood method is discussed in (**5**, **16**, **14**). Until now, the method has not been applied to other than very

small correlation matrices because of the large amount of computations required. From a computational point of view, Lawley's method can be described essentially as equivalent to the task of finding the latent roots and vectors of a modified correlation matrix, the correlations being modified by dividing them by a simple function of the unknown latent vectors.

Extraction of Factors

The application of Lawley's method to the actual data was carried out on Whirlwind I, a high-speed digital electronic computer. The computing program was written by the author with a view to minimizing the use of computer time in case convergence should require hundreds of iterations. A single iteration with this program required roughly 12 seconds, the time varying somewhat with the number (m) of factors hypothesized.

The original hypothesis of the author suggested that m should be at least 9 for the 33-variable matrix analyzed. However, application of Lawley's method to the initial set of trial values for the factor loadings failed because the computations generated imaginary numbers. Extremely close initial approximations to the solution were necessary whenever m was at all large.

The problem was dealt with as follows. Computations were first carried out with $m = 4$. Initial trial values of the factor loadings were arbitrary except that (a) calculated loadings on the first centroid factor were used for the first column, (b) the remaining trial values were selected so that the sum of squares of the trial values for any one variable was equal to the highest correlation with that variable. The iterations were successfully completed for $m = 4$. The resulting estimates of the factor loadings were used as the first four columns of the trial values needed to start the iterations with $m = 5$; the fifth column of these trial values was set up in accordance with informed guesses based on the matrix of residuals. The trial values for $m = 6$ were set up in the same way from the results obtained with $m = 5$, and so forth. In every case after $m = 4$, the initial trial values proved to be close approximations to the corresponding final factor loadings. No further imaginary numbers were encountered.

Convergence of the iterative process was rapid, as shown by the second column of Table 3. The criterion used for stopping iteration required that the largest discrepancy between the corresponding factor loadings produced by two successive iterations should remain less than .002 throughout ten successive iterations.

The matrix of residuals obtained with each value of m was tested for significance by means of Lawley's chi-square test. Information about the progress of the computations and about the chi-square significance tests is given in Table 3. Although arguments could be advanced for extracting an eleventh factor, it was decided to stop with ten.

The orthogonal unrotated matrix of the maximum likelihood estimates

TABLE 3

Tests of Significance and Other Information According to the Number (m) of Factors Hypothesized

Number of Factors Hypothesized (m)	Number of Iterations Required for Convergence	Sum of Latent Roots	Chi-Square Calculated from Residuals	Degrees of Freedom for Chi-Square	Probability Level for Chi-Square
4	35	61	2,605	402	< .01
5	22	69	893	373	< .01
6	23	75	662	345	< .01
7	28	78	530	318	< .01
8	26	80	436	292	< .01
9	25	83	357	267	< .01
10	28	88	284	243	.07

of the factor loadings is given in Table 4. The communality for each test and the latent root for each factor are also shown. Each latent root is the weighted sum of the squares of the loadings on the corresponding factor, the weight for each test being the reciprocal of its uniqueness.

Estimation of Unrotated Factor Loadings for Experimentally Dependent Variables

The six variables in Table 4 with loadings enclosed in parentheses were not included in the 33-variable correlation matrix from which the factors were extracted. These loadings were estimated by the method briefly outlined in the following paragraphs.

The usual factor equation, $R \cong FF'$ ($\cong$ is used to indicate approximate equality), may be written

$$\left[\begin{array}{c|c} P & Q \\ \hline Q' & S \end{array}\right] \cong \left[\begin{array}{c} G \\ \hline H \end{array}\right] \cdot \left[\begin{array}{c|c} G' & H' \end{array}\right] \tag{1}$$

where P is the 33-variable matrix of correlations used for extracting the factors, Q is the matrix of the correlations of these thirty-three variables with the six variables that were omitted from P, S is the matrix of the intercorrelations of these six variables, G is the matrix of the factor loadings of the thirty-three variables and H is the matrix of the factor loadings of the six variables. Assuming that the entire matrix, R, has the same common factors as does P, it follows that G is the matrix of factor loadings obtained by analyzing P. H can then be determined from the equations

$$HG' \cong Q', \tag{2}$$

$$HH' \cong S. \tag{3}$$

TABLE 4

Unrotated Factor Coefficients
(decimal points omitted)

Variable No.	I	II	III	IV	V	VI	VII	VIII	IX	X	Communality
1	34	16	09	-04	17	08	-04	-04	-07	-02	192
2	65	51	-25	-03	-17	18	01	12	10	-02	832
3	56	51	-24	-04	-10	12	12	04	-06	-03	672
4	61	46	-25	-10	-12	08	-05	00	-12	-02	691
5	60	55	-24	-04	-07	14	05	04	12	05	774
6	66	46	-16	-03	13	-10	00	01	-02	03	699
7	68	59	-15	-04	17	-09	04	-02	-04	00	875
8	67	56	-12	00	19	-19	01	-06	01	-06	853
9	(48)	(34)	(11)	(06)	(38)	(-11)	(03)	(-04)	(02)	(-02)	(519)
10	49	-36	-16	11	01	02	-14	05	13	01	446
11	57	-52	-22	-03	-02	11	-02	02	-15	07	689
12	55	-59	-26	-03	02	10	-11	-02	-03	13	765
13	60	-58	-29	-01	-02	-01	-07	-03	-03	12	790
14	60	-57	-26	01	10	01	06	-03	-01	-04	761
15	58	-61	-31	02	11	01	12	-02	02	-12	847
16	60	-61	-24	06	11	01	08	-05	03	-05	826
17	(30)	(-24)	(-12)	(07)	(25)	(-05)	(28)	(-02)	(05)	(-31)	(406)
18	62	02	24	38	-25	02	-04	14	00	01	669
19	59	-07	16	22	-31	-06	-03	03	-10	-12	556
20	59	-07	23	24	-29	-01	04	01	-13	12	580
21	62	-02	17	37	-27	-06	-12	00	-08	-04	642
22	64	-02	23	43	-13	-11	-02	-08	08	02	684
23	63	-05	21	33	-14	-11	02	03	-01	-11	601
24	61	04	24	39	-11	-05	-01	-06	06	-04	604
25	(39)	(04)	(21)	(19)	(16)	(-04)	(11)	(03)	(08)	(-13)	(297)
26	24	12	52	26	39	22	06	-08	06	01	618
27	28	10	55	38	32	21	00	-13	06	09	696
28	26	03	25	05	42	02	-11	13	-12	-02	358
29	36	-09	08	09	48	00	-14	26	-01	-09	483
30	13	04	43	09	49	09	-09	19	-14	-08	531
31	58	39	12	-28	-09	14	-24	-19	-08	-11	718
32	38	04	34	-32	-03	14	-06	-21	01	-11	446
33	64	-41	10	-26	07	-11	-22	08	15	01	754
34	62	-15	48	-35	-12	-05	02	01	03	-04	781
35	60	-20	62	-28	-05	-01	13	04	-01	05	887
36	05	-09	15	-05	-05	-11	-20	-10	15	-09	127
37	(66)	(57)	(-11)	(-03)	(20)	(-07)	(02)	(-02)	(-01)	(00)	(827)
38	(57)	(-58)	(-29)	(03)	(14)	(01)	(14)	(-02)	(02)	(-13)	(807)
39	(64)	(-04)	(23)	(32)	(-07)	(-09)	(02)	(04)	(02)	(-14)	(600)
Latent Roots	40.92	21.90	11.28	5.10	4.09	1.34	1.08	0.83	0.65	0.58	

Since (1) and (2) never hold exactly in practice, (2) represents an inconsistent set of simultaneous linear equations, there being more equations than unknowns. In practice, (3) is totally ignored. A least-squares (but not a maximum likelihood) approximate solution for (2) can be obtained (9) by postmultiplying both sides by $G(G'G)^{-1}$, the result being

$$H = Q'G(G'G)^{-1}. \tag{4}$$

It seemed more appropriate, however, and also computationally easier, in the present case where maximum likelihood procedures had been employed, to post-multiply (2) by $S^{-2}G(G'S^{-2}G)^{-1}$, S^2 being the 33 × 33 diagonal matrix whose elements are the uniquenesses of the 33 variables in P. The result is

$$H = Q'S^{-2}G(G'S^{-2}G)^{-1}. \tag{5}$$

A rigorous justification for (5) is not immediately available. Sufficient justification is apparent, however, when it is pointed out that in Lawley's method of analysis $G'S^{-2}G$ is the diagonal matrix whose elements are the latent roots, and further that the basic equation of Lawley's method can be written

$$G = (P - S^2)S^{-2}G(G'S^{-2}G)^{-1}. \tag{6}$$

The best justification lies in the clarity of the results obtained, as will be seen presently.

Rotation

The rotation of the original factor matrix toward psychologically meaningful oblique factors was carried out on the matrix rotator at The Adjutant General's Office. Extensive final rotations were made by desk calculator. Variables 37, 38, and 39 (scores not "corrected for guessing") were not available during the rotations.

The main guiding principle in all rotations was psychological meaningfulness, as interpreted according to the notions of the writer. The facility with which rotations could be made on the matrix rotator encouraged persistence in the ultimately unsuccessful attempt to find an arithmetic-reasoning speed factor. A total of 497 rotations were carried out, each involving the shift of only one axis.

Table 5 gives the orthogonal projections of the thirty-nine variables on the reference axes—frequently referred to as the "loadings on the rotated factors." Since the term "factor loading" has been used with various meanings in oblique analyses, these projections will hereafter be referred to as "factor coefficients." Table 6 gives the transformation matrix for rotating Table 4 into Table 5. Table 7 gives the intercorrelations among the primary vectors.

TABLE 5

Rotated Factor Coefficients

(decimal points and initial zeros omitted)

		V I	S II	M III	N IV	P V	v VI	s VII	G VIII	H IX	X X
1.	Word Fluency (R)	25	8	-3	6	7	7	0	9	11	-5
2.	Verbal (A)	75	2	3	2	5	-9	0	6	-6	12
3.	Vocabulary (L)	67	1	2	-3	3	3	6	2	2	-8
4.	" (L)	65	4	5	-9	7	8	-4	4	13	-6
5.	" (M)	77	2	-7	9	-5	2	-3	3	-9	10
6.	" (S)	63	4	2	-5	2	27	-3	0	-1	1
7.	" (S)	72	-1	-1	-4	2	30	1	0	2	-3
8.	" (S)	67	-3	0	-1	-5	39	-5	-2	-2	2
9.	" (LIA)	38	1	-3	7	4	28	5	0	0	1
10.	Spatial Relations (A)	4	43	5	6	2	-6	1	-5	-5	20
11.	Intersections (L)	-2	63	5	-3	4	-2	3	2	6	-11
12.	" (L)	-2	71	-4	5	-4	0	-4	-1	2	3
13.	" (M)	-2	67	3	-0	-7	7	-1	-2	-0	2
14.	" (S)	-1	67	0	0	-2	2	20	-1	2	-1
15.	" (S)	-2	69	-1	-2	-1	-3	30	-2	1	1
16.	" (S)	-4	68	0	4	-5	1	23	-3	0	2
17.	" (LIA)	3	29	-2	-4	5	-2	49	-2	2	-3
18.	Mathematics (A)	2	-4	50	6	9	-9	-2	-2	-9	0
19.	Arithmetic Reasoning (L)	-2	-1	50	-9	9	-5	8	7	8	-8
20.	" " (L)	-4	4	45	4	-4	4	-7	6	-3	-17
21.	" " (M)	-1	0	51	1	7	0	-1	-6	7	-3
22.	" " (S)	1	2	40	17	-9	9	3	-8	-5	5
23.	" " (S)	-1	0	46	0	6	2	13	0	0	-2
24.	" " (S)	6	-1	38	16	-4	3	7	-4	-1	4
25.	" " (LIA)	8	1	16	8	5	2	20	4	-5	3
26.	Number Speed (R)	3	1	-3	40	2	-1	6	4	1	-1
27.	" " (R)	-3	2	3	48	-4	2	-3	-4	0	0
28.	Cancellation (R)	4	9	2	0	27	10	-3	2	5	-5
29.	Picture Discrimination (R)	8	23	1	-5	35	2	4	-5	-4	10
30.	Number Checking (R)	-7	-1	5	2	36	0	1	6	6	-8
31.	English (G)	54	-3	2	1	6	2	-8	36	35	1
32.	Foreign Language (G)	20	2	-1	9	-5	-5	5	51	24	0
33.	Eng'g. Draw. & Des. Geom. (G)	6	36	8	-8	5	4	-6	38	-4	23
34.	Chemistry (G)	6	1	26	-8	-1	0	2	67	2	-2
35.	Mathematics (G)	-5	2	28	-1	-4	2	0	66	-7	-12
36.	Conduct (G)	-8	-7	6	2	-3	-1	-2	15	8	21
37.	Vocabulary (NR)	70	-1	-3	0	3	28	0	1	2	1
38.	Intersections (NR)	-1	67	-2	-1	0	-2	32	-2	1	0
39.	Arithmetic Reasoning (NR)	2	2	42	2	8	0	16	2	-1	1

TABLE 6

Transformation Matrix

	I	II	III	IV	V	VI	VII	VIII	IX	X
I	415	324	192	020	009	109	053	181	011	007
II	647	-561	-047	008	030	141	-119	-056	015	003
III	-389	-448	352	129	038	-060	-078	497	000	-086
IV	-308	-036	337	295	035	-031	082	-784	-082	002
V	132	362	-510	161	167	294	091	-192	005	034
VI	273	408	-416	525	112	-684	030	051	172	-020
VII	007	002	-029	056	-350	059	532	090	-392	-432
VIII	-049	-189	357	-497	650	-345	-112	026	-540	089
IX	261	-042	-336	474	-409	-259	081	183	-520	888
X	013	207	-223	346	-493	468	-811	-145	-497	-092

TABLE 7

Correlations between Primary Vectors

		V I	S II	M III	N IV	P V	v VI	s VII	G VIII	H IX	X X
V	I	1.00	.14	.44	.00	-.11	-.08	-.05	-.11	.03	-.13
S	II	.14	1.00	.49	-.08	-.08	.05	-.02	.17	-.01	.20
M	III	.44	.49	1.00	.29	-.04	.13	.03	.13	.09	.19
N	IV	.00	-.08	.29	1.00	.71	.56	.28	.40	-.18	-.09
P	V	-.11	-.08	-.04	.71	1.00	.66	.28	.40	-.31	.01
v	VI	-.08	.05	.13	.56	.66	1.00	.44	.42	-.14	.27
s	VII	-.05	-.02	.03	.28	.28	.44	1.00	.12	-.22	.12
G	VIII	-.11	.17	.13	.40	.40	.42	.12	1.00	-.12	.01
H	IX	.03	-.01	.09	-.18	-.31	-.14	-.22	-.12	1.00	.27
X	X	-.13	.20	.19	-.09	.01	.27	.12	.01	.27	1.00

If the last one or two factors are excluded from consideration, the clarity of the factor structure in Table 5 is made apparent by the visually obvious distinction between 2-digit and 1-digit coefficients. The 1-digit coefficients may be conveniently dismissed as insignificant. *Each 2-digit coefficient without exception has an obvious realistic interpretation.*

In most factor analyses it is customary to ignore coefficients less than .30 or .20, say, as not reliably different from zero. Standard errors for individual factor coefficients have not been computed for the present study; however, with correlations based on 649 cases, as in the present study, the standard error of a correlation coefficient is about .04 for correlations in the neighborhood of zero and about .01 for correlations in the neighborhood of .80. It is to be expected, therefore, that the factor coefficients will have some meaning even in the range from .10 to .20. This will be seen to be actually the case.

Interpretation of Factors

The first three factors of Table 5 correspond to the three aptitude areas about which the present study is centered. They are "level" factors, in contrast to the next four, which are speed factors. The eighth and to a large extent the ninth factors are determined by academic grades. The tenth and last factor seems to have no simple interpretation. All these factors will now be discussed in more detail.

Factor I(V) is the *verbal factor*. In addition to the experimental vocabulary tests, the following variables have two-digit coefficients for this factor, as would be expected:

2.	Verbal Test (A)	.75
31.	English Grade	.54
1.	Word Fluency (R)	.25
32.	Foreign Language Grade	.20

Factor II (S) is a *space factor*. In addition to the experimental intersections tests, the following variables have two-digit coefficients for this factor, as would be expected:

10.	Spatial Relations (A)	.43
33.	Engineering Drawing and Descriptive Geometry Grades	.36
29.	Picture Discrimination (R)	.23

The picture discrimination test is a reference test for the perceptual-speed factor, but the test obviously requires also the ability to perceive and discriminate spatial patterns.

Factor III (M) is a *mathematical-reasoning factor*. In addition to the experimental arithmetic-reasoning tests, the following variables have two-digit coefficients for this factor, as would be expected:

18.	Mathematics (A)	.50
35.	Mathematics Grade	.28
34.	Chemistry Grade	.26

Factor IV (N) is the *number-speed factor*, determined by the two reference tests included for this purpose. The only other variables with two-digit coefficients for this factor are two of the speeded arithmetic-reasoning tests:

22, 24.	Arithmetic Reasoning (speeded)	.17, .16

Factor V (P) is the *perceptual-speed factor*, determined by the three reference tests included for this purpose. No other variables have two-digit coefficients for this factor.

Factor VI (v) is clearly the *verbal-speed factor* that the present analysis

was designed to isolate (if it actually existed) and to study. All the two-digit coefficients for this factor are listed below:

8, 7, 6. Vocabulary (speeded)	.39, .30, .27
9. Vocabulary (last item attempted)	.28
37. Vocabulary (speeded; number-right score)	.28
28. Cancellation (R)	.10

The cancellation test is a reference test for the perceptual-speed factor. The coefficient of .10 for this test on the verbal-speed factor is not large enough to be of interest; a positive coefficient might be expected, however, in view of the fact that this test requires rapid work with alphabetical and verbal symbols.

Factor VII (s) is clearly the *spatial-speed factor* that the present study was designed to isolate (if it actually existed) and to study. All the two-digit coefficients for this factor are listed below:

17. Intersections (last item attempted)	.49
38. Intersections (speeded; number-right score)	.32
15, 16, 14. Intersections (speeded)	.30, .23, .20
25. Arithmetic Reasoning (last item attempted)	.20
39. Arithmetic Reasoning (speeded; number-right score)	.16
23. Arithmetic Reasoning (speeded)	.13

The fact that all of the speed scores on the arithmetic-reasoning tests have small positive loadings on the spatial-speed factor is consistent with the fact that the arithmetic-reasoning tests contain a considerable proportion of simple geometry and other items that involve graphic illustrations, these being printed in the test booklets alongside the items.

Factor VIII (G) is an *academic-grades factor*. No variables other than the six academic grades have two-digit coefficients for this factor.

Factor IX (H) appears to be some sort of verbal-academic-grade factor, as indicated by its two-digit coefficients, which are as follows:

31. English Grade	.35
32. Foreign Language Grade	.24
4. Vocabulary (level)	.13
1. Word Fluency (R)	.11

Factor X (X) does not suggest any ready interpretation.

The Correlations among Factors

The correlations among the primary vectors in Table 7 are of paramount interest. First, it should be pointed out that the reference axis for the verbal factor was arbitrarily set approximately orthogonal to the reference axis for the verbal-speed factor, since it was felt that interpretation would be hindered

by a choice of reference axes that would give the speeded verbal tests loadings of zero on the verbal factor. For the same reason, the spatial-factor axis was set roughly orthogonal to the spatial-speed axis. In each of these cases, the correlations between the speeded tests and the corresponding level factor are therefore represented approximately by the factor coefficients of the speeded tests and not by the corresponding near-zero correlation in Table 7.

Because of the considerable indeterminacy as to the proper position of the primary vector for the ninth factor, the verbal-factor axis and the academic-grades axis were both set approximately orthogonal to the axis for the ninth factor.

The mathematical-reasoning factor shows correlations of .44 and .49 with the verbal and spatial factors, respectively. These correlations are reasonable in view of the fact that the arithmetic-reasoning tests include verbally presented problems, geometry problems, and other graphically presented problems. The only other correlations in Table 7 as large as these are between various speed factors. In fact, the main thing about Table 7 is the consistently positive intercorrelations of the four speed factors that have been isolated. In general, these correlate much more highly with each other than they do with the three "level" factors, thus demonstrating the existence of a second-order general speed factor.

The Relation of Grades to Speed

The academic-grade factor is seen from Table 7 to be positively correlated with all four of the speed factors. The ninth factor, however, which is determined mainly by grades in English and in Foreign Language, has negative correlations with each of the four speed tests and with the academic-grade factor itself. In order to interpret the relation of course grades to the various speed factors, it is necessary to obtain the actual correlations of each of the grades with the primary vectors for each of the speed factors.

As shown in Table 8, each of the course grades, with one minor exception, is positively correlated with each of the four speed factors. Although these relationships are not high, there is clear evidence of a positive relation between grades at Annapolis and speed.

TABLE 8

Correlations between Course Grades and Primary Vectors

	V I	S II	M III	N IV	P V	V VI	S VII	G VIII	H IX	X X
31. English (G)	61	12	39	20	10	15	-10	36	35	41
32. Foreign Language (G)	18	13	24	26	12	17	04	54	20	08
33. Eng'g. Draw. & Des. Geom. (G)	09	68	46	09	15	30	02	54	-00	-05
34. Chemistry (G)	17	36	51	28	16	28	07	76	-02	04
35. Mathematics (G)	08	35	51	43	27	35	11	84	-15	02

Discussion

No speed factor for the arithmetic-reasoning tests could be isolated, although the attempt was persistently made. The factor coefficients indicate that the speeded arithmetic-reasoning tests tend to involve the number-speed factor, the verbal-speed factor, and the spatial-speed factor to a slightly greater extent than do the unspeeded tests, as might reasonably be expected. The picture is somewhat confused by the fact that test 23 behaves slightly differently from the parallel tests 22 and 24, and test 19, from the parallel test 20. It may be that an arithmetic-reasoning factor exists in the data but is so very unimportant that it was not separated from "noise" in the analysis.

The results obtained for the last-item-attempted scores are of particular interest. It will be remembered that these scores were not used to determine the common-factor space, since it was desired to center the present study primarily around the type of scores normally used for aptitude tests. It is nevertheless found, for both the verbal and the spatial tests, that the LIA score is a purer measure of the corresponding speed factor than are the corrected-for-guessing scores on any of the three speeded tests.

It is noteworthy that in all three cases the "moderately" speeded tests (M) are like the level tests and not like the speeded tests, even though only 50 to 75 per cent of the examinees responded to the last item.

Variables 37, 38, and 39—the number-right scores corresponding to variables 7, 15, and 23—have loadings so similar to the "corrected-for-guessing" scores on the same tests as to be virtually indistinguishable from the latter.

With the exception of English, the academic grades all have higher loadings on the academic-grade factor than they do on any of the aptitude factors. This situation clearly shows that the course grades have a reliable, and therefore theoretically predictable, variance over and above that actually predicted by the aptitude tests. Whether this variance is attributable to personality factors or to other causes cannot be determined from the present study.

Summary and Conclusions

The present study was designed to investigate the existence and inter-relations of various speed factors, and their relation to academic course grades.

Speeded and unspeeded, but otherwise parallel, tests of vocabulary, spatial ability, and arithmetic reasoning were administered to 649 entering students at the U. S. Naval Academy at Annapolis. Also included in the factorial analysis were scores on certain regular admissions examinations, scores on certain specially prepared reference tests, and end-of-year course grades at Annapolis.

Extraction of factors from the 33-variable correlation matrix was carried out by Lawley's maximum likelihood method, the calculations being done on the Whirlwind, a high-speed electronic computer. Factoring was continued until, after the extraction of the tenth factor, a significance test on the matrix of residuals showed them to be no longer statistically significant.

Rotation to psychologically meaningful oblique axes was carried out with the help of the matrix rotator at The Adjutant General's Office. The tenth rotated factor was found to be difficult or impossible to interpret. With this exception, the structure of the factor matrix was found to be so clear that a ready interpretation existed for every factor coefficient above .09.

As would be expected, three of the factors obtained were verbal, spatial, and mathematical-reasoning factors. The reference tests included in the battery yielded the expected number-speed factor (ordinarily called simply the number factor) and perceptual-speed factor. The academic grades in the battery were found to define not only a general academic-grade but also a verbal-academic-grade factor. Finally, a verbal-speed and a spatial-speed factor were clearly identified and distinguished from the number-speed and the perceptual-speed factors. No arithmetic-reasoning speed factor was isolated.

The primary vectors for all four speed factors were found to be positively intercorrelated, demonstrating the existence of a general speed factor at the second-order level.

All correlations between course grades and the four speed factors, with one small exception, were found to be positive, although not large. It is to be concluded that speed of various kinds plays some part in the course grades studied, and that speededness in the admissions examinations is to this extent justified. It would seem that tests on which 50 to 75 per cent of the examinees reach the last item do not involve the speed factors needed; apparently, only very highly speeded tests involve these factors to any appreciable extent.

REFERENCES

1. Bartlett, M. S. Tests of significance in factor analysis. *Brit. J. Psychol., statist. Sect.*, 1950, **3**, 77–85.
2. Bartlett, M. S. A further note on tests of significance in factor analysis. *Brit. J. Psychol., statist. Sect.*, 1951, **4**, 1–2.
3. Davidson, W. M., and Carroll, J. B. Speed and level components in time-limit scores: a factor analysis. *Educ. psychol. Measmt.*, 1945, **5**, 411–427.
4. Dubois, P. H. A speed factor in mental tests. *Arch. Psychol.*, 1932, No. 141, pp. 38.
5. Emmett, W. G. Factor analysis by Lawley's method of maximum likelihood. *Brit. J. Psychol., statist. Sect.*, 1949, **2**, 90–97.
6. French, J. W. The description of aptitude and achievement tests in terms of rotated factors. *Psychometric Monogr.* No. 5, 1951.
7. French, J. W. Selected tests for reference vectors. Princeton: Educational Testing Service, 1953.

8. Goheen, H. W., and Kavruck, S. Selected references on test construction, mental test theory, and statistics, 1929–1949. Washington, D. C. Superintendent of Documents, 1950.
9. Harris, C. W. The symmetrical idempotent matrix in factor analysis. *J. exper. Educ.*, 1951, **19**, 238–246.
10. Henrysson, S. The significance of factor loadings—Lawley's test examined by artificial samples. *Brit. J. Psychol., statist. Sect.*, 1950, **3**, 159–166.
11. Kendall, M. G. Factor analysis. Part I: Factor analysis as a statistical technique. *J. roy. statist. Soc., Series B*, 1950, **12**, 60–73.
12. Lawley, D. N. The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc. Edinburgh*, 1940, **60**, 64–82.
13. Lawley, D. N. Further investigations in factor estimation. *Proc. Roy. Soc. Edinburgh*, 1941, **61**, 176–185.
14. Lawley, D. N. The application of the maximum likelihood method to factor analysis. *Brit. J. Psychol.*, 1943, **33**, 172–175.
15. Lawley, D. N. Problems in factor analysis. *Proc. Roy. Soc. Edinburgh, Section A*, 1949, **62**, 1–6.
16. Lawley, D. N. The maximum likelihood method of estimating factor loadings. In Thomson, G. H. (Ed.): The factorial analysis of human ability. (5th Ed.) London: Univ. of London Press, 1951. Pp. 321–329.
17. Mahan, W. W., Jr. The factorial comparison of five mental abilities using scores derived from speeded and unspeeded tests. Unpublished doctor's dissertation, Univ. of Texas, 1953.
18. Myers, C. T. The factorial composition and validity of differently speeded tests. *Psychometrika*, 1952, **17**, 347–352.
19. Rao, C. R. Estimation and tests of significance in factor analysis. Urbana: Univ. of Illinois. (Mimeo.)
20. Rimoldi, H. J. A. Personal tempo. *J. abn. soc. Psychol.*, 1951, **46**, 283–303.
21. Sutherland, J. D. The speed factor in intelligent reactions. *Brit. J. Psychol.*, 1934, **24**, 276–294.
22. Tate, M. W. Individual differences in speed of response in mental test materials of varying degrees of difficulty. *Educ. psychol. Measmt*, 1948, **8**, 353–374.
23. Tucker, L. R. A restatement of the equations for Lawley's maximum likelihood method of estimating factor loadings. Princeton: Princeton Univ., 1953. (Mimeo.)
24. Vernon, P. E. The structure of human abilities. New York: Wiley, 1950.
25. Whittle, P. On principal components and least square methods of factor analysis. *Skandinavisk Aktuarietidskrift*, 1952, **35**, 223–239.
26. Uppsala symposium on psychological factor analysis, 17–19 March 1953. *Nordisk Psykologi*, 1953, Monograph Series No. 3.

Manuscript received 11/29/54
Revised manuscript received 3/9/55

PSYCHOMETRIKA—VOL. 21, NO. 1
MARCH, 1956

OPTIMAL TEST LENGTH FOR MAXIMUM DIFFERENTIAL PREDICTION*

PAUL HORST

UNIVERSITY OF WASHINGTON

For the case of a single criterion a method is already available for determining the optimal distribution of testing time for a battery of predictors, assuming that intercorrelation, validity, and reliability data are available for predictors of arbitrary lengths. In this article a modification and generalization of the method is presented for the case of differential prediction involving a number of criterion variables. A numerical example is given to illustrate the method, after which the mathematical rationale is outlined.

I. *The Problem*

In (2) the importance of techniques for predicting success differentially in each of a number of different activities from a single battery of predictors was discussed. It was assumed that intercorrelations for a large battery of predictor variables were available and also correlations between these predictors and a large number of criterion variables. The problem was to select from this larger battery of predictors that subset of specified size which would yield the maximum index of differential prediction for the criterion variables. The index of differential prediction efficiency was taken to be a simple function of the average of the variances for the predicted difference scores for all possible pairs of criterion variables. The larger this average variance the greater the differential prediction efficiency of the battery. It was shown that this index is equivalent to the difference between the average variance of the predicted criterion measures and the average of their covariances, assuming standard measures for both predictors and criteria, and that the predicted criteria are the "least squares" estimates. A method for selecting that subset of predictors of specified size which would yield the maximum index of differential prediction was presented.

The method referred to tacitly assumes that all predictors in the battery take the same amount of administration time, so that all subsets of the same size would also take the same amount of administration time. Usually this will not be the case. A more general approach to the problem might be to start with a given battery of predictor variables and inquire how the ad-

*This research was carried out under Contract Nonr-477(08) between the University of Washington and the Office of Naval Research. Most of the computations were carried out by Robert Dear, Charlotte MacEwan, and Donald Mills. Much credit is due the typist, Elizabeth Cross. Supervision of both computational and editorial activities was provided by William Clemans. To each of these able contributors I am deeply grateful.

ministration time for each predictor should be altered so that for a specified over-all testing time the index of differential prediction efficiency will be a maximum. This approach would allow for increasing the length of an experimental battery as well as for decreasing it.

As a matter of fact, for the case of a single criterion a method is already available (1) for determining the optimal distribution of testing time for a battery of predictors, assuming that intercorrelation, validity, and reliability data are available for predictors of arbitrary lengths. It is the purpose of this article to present a modification and generalization of the method for the case of differential prediction involving a number of criterion variables.

In this presentation testing time is taken to be the time actually allotted the examinee for taking the test. A more complete analysis must also take into account the time for reading instructions, practice exercises, passing out and collecting papers, etc. The method will first be described and illustrated by a numerical example, after which the mathematical rationale will be presented.

II. *Numerical Example*

The predictor variables used in this example are:

(1) Guilford-Zimmerman Aptitude Survey, Part I, Verbal Comprehension
(2) Guilford-Zimmerman Aptitude Survey, Part III, Numerical Operations
(3) Guilford-Zimmerman Aptitude Survey, Part VII, Mechanical Knowledge
(4) A. C. E. Psychological Examination, Quantitative Reasoning
(5) A. C. E. Psychological Examination, Linguistic Reasoning
(6) Cooperative English Test (Form OM), Usage

The matrix of test intercorrelations with reliabilities in the diagonal is given in Table 1. The criterion variables are grade-point averages in each of ten college subjects. The matrix of validity coefficients is given in Table 2.

The over-all testing time for the tests of arbitrary length is 142 minutes. We assume that this time is to be cut in half so that the over-all testing time is 71 minutes. The problem is to determine the time to be allotted to each test so as to maximize the index of differential prediction efficiency.

The traditional assumptions are used here as in (1) with respect to the effect of test length on correlation and will not be repeated. Altering the administration time for any test will, of course, alter the number of items. In the following discussion, the terms test length and testing time are used synonymously.

The method of solution for the new test lengths involves a series of successive approximations. For large numbers of predictor and criterion

TABLE 1

R Matrix of Predictor Intercorrelations with Reliabilities Substituted for Unities in the Diagonal:

$$R = r\text{-}D_u$$

		1	2	3	4	5	6	Σ
1	G-Z 1	.920	.159	.152	.281	.763	.515	2.790
2	G-Z 3	.159	.920	.003	.369	.292	.243	1.986
3	G-Z 7	.152	.003	.920	.200	.142	-.150	1.267
4	ACE-Q	.281	.369	.200	.820	.549	.426	2.645
5	ACE-L	.763	.292	.142	.549	.830	.628	3.204
6	English	.515	.243	-.150	.426	.628	.860	2.522
Σ		2.790	1.986	1.267	2.645	3.204	2.522	14.414

TABLE 2

The r_c' Matrix of Validity Coefficients

	1 G-Z 1	2 G-Z 3	3 G-Z 7	4 ACE-Q	5 ACE-L	6 English	Σ
1 Anthropology	.370	.177	.091	.294	.341	.357	1.630
2 Chemistry	.317	.274	.016	.309	.364	.399	1.679
3 Economics	.339	.211	.008	.241	.334	.323	1.456
4 English	.526	.247	-.075	.262	.488	.524	1.972
5 Foreign Lang.	.295	.287	-.156	.200	.232	.426	1.284
6 Geology	.184	.140	.094	.170	.229	.214	1.031
7 History	.379	.169	-.001	.182	.373	.336	1.438
8 Mathematics	.287	.348	-.088	.350	.336	.401	1.634
9 Psychology	.440	.170	.096	.285	.409	.403	1.803
10 Zoology	.336	.216	.031	.318	.345	.351	1.597
Σ	3.473	2.239	.016	2.611	3.451	3.734	15.524
Σ/10	.347	.224	.002	.261	.345	.373	1.552

variables the solution may become very laborious. It is probable that the solution would be greatly expedited by the use of high-speed computing equipment. Further research may yield more efficient computational procedures.

1. The first computational step is to calculate a matrix α_c' from the matrix r_c' in Table 2. The elements α_c' of Table 3 are the corresponding elements of Table 2 with their column means subtracted. Hence the columns of Table 3 all add to zero.

2. The next step is to compute the elements for a diagonal matrix, Δ. The ith element is the product of the original length of test i multiplied by one minus its reliability. The elements for Δ are given in row 4, labeled $1'$ Δ in Table 4. For the first element we have $\Delta_1 = 25(1.00 - .92) = 2.000$.

3. A first approximation is now required for the altered test lengths. We assume the new test lengths to be proportional to the original test lengths. Therefore, as a first approximation to the new test lengths we take one half

TABLE 3

The α'_c Matrix: Validity Coefficients Expressed in Deviation Form for Each Test

	1	2	3	4	5	6
1	.023	-.047	.089	.033	-.004	-.016
2	-.030	.050	.014	.048	.019	.026
3	-.008	-.013	.006	-.020	-.011	-.050
4	.179	.023	-.077	.001	.143	.151
5	-.052	.063	-.158	-.061	-.113	.053
6	-.163	-.084	.092	-.091	-.116	-.159
7	.032	-.055	-.003	-.079	.028	-.037
8	-.060	.124	-.090	.089	-.009	.028
9	.093	-.054	.094	.024	.064	.030
10	-.011	-.008	.029	.057	.000	-.022
Ck	.003	-.001	-.004	.001	.001	.004
Σ	.003	-.001	-.004	.001	.001	.004

TABLE 4

Computation of $1'D_{b_i}^{-1}$ and $1'\Delta D_{b_i}^{-1}$

First approximation: $1'D_{b_1} = \left(\frac{T_1}{1'D_a 1}\right) 1'D_a$

		1	2	3	4	5	6	Ck	Σ
1	$1'D_a$	25.0	9.0	30.0	23.0	15.0	40.0		142.0
2	$1'D_{b_1}=.5\ 1'D_a$	12.5	4.5	15.0	11.5	7.5	20.0	71.0	71.0
3	$1'D_{b_1}^{-1}$	.080	.222	.0667	.087	.1333	.050		
4	$1'\Delta$	2.000	.720	2.400	4.140	2.550	5.600		17.410
5	$1'\Delta D_{b_1}^{-1}$	.160	.160	.160	.360	.340	.280		1.460

the original test lengths. Row 1 in Table 4 gives the original test lengths. Row 2 of the same table is one half the first.

4. Calculate the reciprocals of the D_b elements. These are given in row 3 of Table 4.

5. Calculate the product of each Δ value in row 4 of Table 4 by the corresponding value immediately above it. The products are entered in row 5 of Table 4. For example, the first value is .160 = 2.000 × .080.

6. Next the elements calculated in step 5 are added to the corresponding diagonal elements of Table 1, and the table is copied into the upper left quadrant of Table 5. The first diagonal element is 1.080 = .160 + .920. Note that the elements below the diagonal are not copied in. The upper right section of Table 5 is α_c, the transpose of Table 3.

TABLE 5

Computation of $(R+\Delta D_{b_1}^{-1})^{-1} a_c = L_1$ Forward Solution

	1A	2A	3A	4A	5A	6A	1B	2B	3B	4B	5B	6B	7B	8B	9B	10B	Check	Σ
1A	1.080	.159	.152	.281	.763	.515	.023	-.030	-.008	.179	-.052	-.163	.032	-.060	.093	-.011	2.953	
2A		1.080	.003	.369	.292	.243	-.047	.050	-.013	.023	.063	-.084	-.055	.124	-.054	-.008	2.145	
3A			1.080	.200	.142	-.150	.089	.014	.006	-.077	-.158	.092	-.003	-.090	.094	.029	1.423	
4A				1.180	.549	.426	.033	.048	-.020	.001	-.061	-.091	-.079	.089	.024	.057	3.006	
5A					1.170	.628	-.004	.019	-.011	.143	-.113	-.116	.028	-.009	.064	.000	3.545	
6A						1.140	-.016	.026	-.050	.151	.053	-.159	-.037	.028	.030	-.022	2.806	
	2.950	2.146	1.427	3.005	3.544	2.802	.078	.127	-.096	.420	-.268	-.521	-.114	.082	.251	.045	15.878	
.9259 1	1.080	.159	.152	.281	.763	.515	.023	-.030	-.008	.179	-.052	-.163	.032	-.060	.093	-.011	2.953	2.953
.9461 2		1.057	-.019	.328	.180	.167	-.050	.054	-.012	-.003	.071	-.061	-.060	.133	-.068	-.006	1.711	1.711
.9452 3			1.058	.166	.038	-.220	.085	.019	.007	-.102	-.149	.114	-.009	-.079	.080	.030	1.037	1.038
1.0215 4				.979	.289	.275	.029	.036	-.015	-.029	-.046	-.048	-.067	.076	.008	.057	1.545	1.544
1.9455 5					.514	.163	-.023	.020	.001	.029	-.069	.020	.036	-.009	.005	-.009	.676	.678
1.4430 6						.693	-.002	.019	-.039	.044	.070	-.041	-.037	.001	.009	-.023	.694	.694

7. We next calculate a matrix L_1 by premultiplying the matrix α_e by the inverse of the matrix in the upper left quadrant of Table 5. The computations for the forward solution are given in the two lower quadrants of Table 5 and in Table 6. The back solution is given in Table 7. The procedure for multiplying a matrix by the inverse of a symmetric matrix is outlined in (3).

8. The second approximation to the new test lengths is computed in the lower section of Table 7 as follows:

Row a consists of the sum of squares of column elements of the L_1 matrix. For example, the first element in row a, namely, .0626, is the sum of squares of the first ten elements in column 1 of Table 7.

Row b is copied from row 4 of Table 4.

Row c consists of the products of corresponding elements in the two preceding lines. For example, $.1251 = .0626 \times 2.00$.

Row d consists of the square roots of corresponding entries in the preceding line. For example, $.3537 = \sqrt{.1251}$. The computations to the right of this line and designated s are obtained by dividing the over-all new testing time, 71 minutes, by 1.8823, the sum of the elements in the row. This gives $s = 37.7198$.

Row e is a check row. Each element in the second line above it is divided by the element immediately above. Thus $.1251/.3537 = .3537$.

Row f is obtained by multiplying each element in row d by s. For example, the first element is $13.3415 = .3537 \times 37.7198$. This line gives the second approximation to the new test lengths.

Row g is obtained by dividing each element in the preceding row into the corresponding value in row b. For example, the first value is $.150 = 2.00/13.3415$.

Row h is a check on the preceding row. Each element in row g is multiplied by the corresponding element in row f to give row h, which should correspond, within limits of rounding errors, to row b. For example, the first element is $2.001 = .150 \times 13.3415$.

Row i is obtained by adding the elements in row g to the corresponding reliabilities. For example, the first element is $1.070 = .150 + .920$.

9. A new L_2 matrix is now computed by repeating steps 6 and 7 and using the elements of row i of Table 7 in the diagonal positions of Table 1. The new L_2 matrix is given in transposed form in Table 8, rows 1 through 10.

10. Step 8 is repeated in rows a through i of Table 8. Row f of this table gives a third approximation to the altered test lengths.

Steps 6, 7 and 8 are repeated to get successive approximations to the test lengths. The calculations were carried to 5 successive approximations for the new test lengths, not counting the first. These are summarized in Table 9. As will be seen, the iterations have not completely stabilized. However, for practical purposes, the approximation is doubtless adequate.

TABLE 6

Computation of $(R+\Delta D_{b_1}^{-1})^{-1} \alpha_c = L_1$ continued

	1	2	3	4	5	6	1	2	3	4	5	6	7	8	9	10	Check	Σ
1	1.000																	
2		1.000																
3			1.000															
4				1.000														
5					1.000													
6						1.000												
1	-1.000	-.147	-.141	-.260	-.706	-.477	-.021	.028	.007	-.166	.048	.151	-.030	.056	-.086	.010	-2.734	-2.734
2		-1.000	.018	-.310	-.170	-.158	.047	-.051	.011	.003	-.067	.057	.057	-.126	.064	.006	-1.619	-1.619
3			-1.000	-.157	-.036	.208	-.080	-.018	-.007	.096	.141	-.108	.009	.075	-.076	-.028	- .981	- .981
4				-1.000	-.295	-.281	-.030	-.037	.015	.030	.047	.049	.068	-.077	-.008	-.058	-1.577	-1.577
5					-1.000	-.317	.045	-.039	-.002	-.056	.134	-.039	-.070	.018	-.010	.018	-1.319	-1.318
6						-1.000	.003	-.027	.056	-.063	-.101	.059	.053	-.001	-.013	.033	-1.001	-1.001

TABLE 7

Computation of $(R+\Delta D_{b_1}^{-1})^{-1} a_c = L_1$ Back Solution

L_1' Matrix

		1	2	3	4	5	6	1	2	3	4	5	6	7	8	9	10	Check		$\lvert L_1 \rvert 1$
1		.039	-.051	.074	.044	-.044	-.003	-1										0	.001	.255
2		-.075	.035	.019	.021	.030	.027		-1									0	-.001	.207
3		.008	-.004	-.005	-.005	.020	-.056			-1								0	.003	.098
4		.136	-.002	-.075	-.058	.036	.063				-1							0	.000	.370
5		.031	.085	-.110	-.026	-.166	.101					-1						0	.001	.519
6		-.159	-.040	.101	-.050	.058	-.059						-1					0	.000	.467
7		.022	-.039	-.011	-.079	.087	-.053							-1				0	.002	.291
8		-.068	.102	-.087	.082	-.018	.001								-1			0	.002	.358
9		.074	-.067	.078	.003	.006	.013									-1		0	.002	.241
10		-.005	-.021	.010	.070	-.008	-.033										-1	0	-.001	.147
a	$1' D_{L_1 L_1'}$	.0626	.0295	.0477	.0268	.0434	.0256													
b	$1'\Delta$	2.00	.72	2.40	4.14	2.55	5.60	Ck 17.41			Σ 17.41									
c	$1' D_{L_1 L_1'} \Delta$	.1251	.0214	.1144	.1110	.1108	.1433													
d	$1'(D_{L_1 L_1'}\Delta)^{\frac{1}{2}}$	.3537	.1458	.3382	.3332	.3328	.3786				1.8823	$s = \dfrac{T_1}{1'(D_{L_1 L_1'}\Delta)^{\frac{1}{2}} 1} = \dfrac{71}{1.8823} = 37.7198$								
e	Ck: $1' D_{L_1 L_1'} \Delta (D_{L_1 L_1'} \Delta)^{-\frac{1}{2}}$	.3537	.1458	.3382	.3332	.3329	.3786													
f	$1' D_{b_2} = 1'(D_{L_1 L_1'}\Delta)^{\frac{1}{2}} s$	13.3415	5.4995	12.7568	12.5682	12.5531	14.2807				71.0000									
g	$1'\Delta D_{b_2}^{-1}$	.150	.131	.188	.329	.203	.392				1.393									
h	Ck: $1' D_{b_2} \Delta D_{b_2}^{-1}$	2.001	.720	2.398	4.135	2.548	5.598	6.663			6.663 = 1' Diag. of $(R+\Delta D_{b_2}^{-1})$									
i	$1' D_{r_{ii}} + 1'\Delta D_{b_2}^{-1}$	1.070	1.051	1.108	1.149	1.033	1.252													

TABLE 8

Computation of $(R+\Delta D_{b_2}^{-1})^{-1} a_c = L_2$ Back Solution

L_2' Matrix

		1	2	3	4	5	6
1		.051	-.053	.072	.052	-.065	.000
2		-.083	.036	.017	.018	.045	.022
3		.003	-.005	-.003	-.009	.024	-.049
4		.132	-.003	-.076	-.062	.053	.052
5		.070	.093	-.108	-.009	-.229	.100
6		-.176	-.043	.100	-.057	.081	-.056
7		.002	-.041	-.009	-.092	.119	-.052
8		-.066	.105	-.085	.086	-.025	.003
9		.076	-.069	.075	.005	.007	.010
10		-.005	-.022	.012	.072	-.012	-.028
a	$1' D_{L_2L_2'}$	.0730	.0326	.0460	.0314	.0836	.0223
b	$1'\Delta$	2.0	.72	2.40	4.14	2.55	5.60
c	$1' D_{L_2L_2'}\Delta$	.1459	.0235	.1104	.1298	.2132	.1250
d	$1'(D_{L_2L_2'}\Delta)^{\frac{1}{2}}$	.3820	.1532	.3323	.3603	.4618	.3536
e	Ck: $1' D_{L_2L_2'}\Delta(D_{L_2L_2'}\Delta)^{-\frac{1}{2}}$	.3820	.1532	.3322	.3602	.4617	.3535
f	$1' D_{b_3} = 1'(D_{L_2L_2'}\Delta)^{\frac{1}{2}} s$	13.2743	5.3236	11.5472	12.5202	16.0473	12.2874
g	$1'\Delta D_{b_3}^{-1}$	.151	.135	.208	.331	.159	.456
h	Ck: $1' D_{b_3}\Delta D_{b_3}^{-1}$	2.004	.720	2.402	4.144	2.552	5.603
i	$1' D_{r_{ii}} + 1'\Delta D_{b_3}^{-1}$	1.071	1.055	1.128	1.151	.989	1.316

$$s = \frac{T_1}{1'(D_{L_2L_2'}\Delta)^{\frac{1}{2}}1} = \frac{71}{2.0432} = 34.7494$$

	Ck	Σ
		2.0432
		-
	71.000	71.000
		1.440
	6.710	6.710 = $1'$ Diag. of $(R+\Delta D_{b_3}^{-1})$

TABLE 9

Successive Approximations to $1'D_b$, for $T_1 = \frac{1}{2}T_0 = \frac{142}{2} = 71$

Approx'n		1	2	3	4	5	6	Σ
$(.5)1'D_a$:	1	12.50	4.50	15.00	11.50	7.50	20.00	71.00
	2	13.34	5.50	12.76	12.57	12.55	14.28	71.00
	3	13.27	5.32	11.55	12.52	16.05	12.29	71.00
	4	13.23	5.20	10.98	12.47	17.62	11.51	71.01
	5	13.31	5.15	10.76	12.46	18.13	11.19	71.00
	6	13.35	5.12	10.70	12.46	18.37	11.00	71.00

Value of ϕ for Successive Values of L

L_1	.227
L_2	.234
L_3	.235
L_4	.236
L_5	.237

11. To compute the successive indices of differential prediction efficiency ϕ_e, we proceed as follows:

(*a*) For the index corresponding to the first approximation to the new test length multiply each element in the L_1 matrix in Table 7 by the corresponding element of Table 3 and sum the products. This is the first entry, .227, in the ϕ column at the right of Table 9.

(*b*) To get ϕ_2 follow the same procedure except use the L_2 matrix in Table 8 instead of L_1 in Table 7.

(*c*) In the same way calculate subsequent ϕ's by using the elements in the corresponding L matrix and the elements in Table 3.

It will be noted that ϕ does not increase much in this particular illustration. It goes from .227, taking the test lengths as one half their original length, to .237 as they approach optimal length. This is an increase of less than 5 per cent even though several of the test lengths are changed greatly. For example, test 5 increases from 7 to 18 minutes while test 6 reduces from 20 to 11 minutes.

The technique was applied to the same data assuming that the total administration time was to be the same as in the original administration, namely, 142 minutes and also assuming it was to be doubled to 284 minutes. Only three approximations to the optimal test lengths were calculated for each of these two conditions. Tables 10 and 11 summarize the results for the two conditions, respectively. The last column in each table shows the index of differential prediction efficiency, ϕ, corresponding to each approximation to optimal test lengths. In both cases the improvement of ϕ as the tests approach optimal length is appreciably greater than for the case of one half the original testing time. This rate of improvement is greatest for double testing time. As can be seen from the right-hand column of Table 11 it goes from .305 to .337, which is approximately a 10 per cent increase. Further research is needed to determine the sensitivity of ϕ to alterations in relative testing time for each of the tests and to variations in total testing time.

TABLE 10

Successive Approximations to $1'D_b$, for $T_1 = T_0 = 142$

Approx'n		1	2	3	4	5	6	Σ	Value of ∅ for Successive Values of L	
(1) $1'D_a$:	1	25.00	9.00	30.00	23.00	15.00	40.00	142.00		
	2	25.38	8.85	20.04	25.55	33.62	28.56	142.00	L_1	.265
	3	25.83	7.60	16.02	25.41	42.43	24.73	142.02	L_2	.282
	4	26.48	7.30	14.72	25.22	44.54	23.73	141.99	L_3	.283

TABLE 11

Successive Approximations to $1'D_b$, for $T_1 = 2T_0 = 2(142) = 284$

Approx'n		1	2	3	4	5	6	Σ	Value of ∅ for Successive Value of L	
(2) $1'D_a$:	1	50.00	18.00	60.00	46.00	30.00	80.00	284.00		
	2	51.51	13.62	30.14	51.31	83.57	53.84	284.00	L_1	.305
	3	54.84	10.84	22.51	51.17	97.72	46.91	284.00	L_2	.334
	4	56.32	10.53	21.24	50.94	99.64	45.34	284.01	L_3	.337

III. *Mathematical Derivation*

In (1) a procedure is developed for altering test lengths in a battery to give maximum multiple correlation with a single criterion. The development of this procedure will be reviewed and the procedure will be extended to the problem of differential prediction. Let

M = the number of cases,
n = the number of predictors,
Z = an $(M \times n)$ matrix of test scores in a battery of altered lengths with the elements of Z of the form $(z_{ij} - \bar{z}_j)/(\sqrt{M}\sigma_{z_j})$,
W = an $(M \times 1)$ vector of criterion scores with elements of the form $(w - \bar{w})/(\sqrt{M}\,\sigma_w)$,
B = an $(n \times 1)$ vector of regression coefficients for estimating W from Z,
r = an $(n \times n)$ matrix of intercorrelations of tests of original lengths,
ρ = an $(n \times n)$ matrix of intercorrelations of tests of altered lengths,
r_c = an $(n \times 1)$ vector of validity coefficients for the tests of original lengths,
ρ_c = an $(n \times 1)$ vector of validity coefficients for the tests of altered lengths,
D_a = an $(n \times n)$ diagonal matrix of original test lengths,

D_b = an $(n \times n)$ diagonal matrix of altered test lengths,
D_e = $D_b D_a^{-1}$ be the ratio of altered to original test lengths,
D_{rii} = the $(n \times n)$ diagonal matrix of reliability coefficients for the tests of original lengths.

Let

$$\delta = [I + (D_e - I)D_{rii}]D_e^{-1}. \tag{1}$$

Let

$$\epsilon = (ZB - W). \tag{2}$$

We wish to minimize $\epsilon'\epsilon$ with the constraining condition $1'D_b1 = T$, where T is the total testing time specified for tests with altered lengths, and 1 is a column vector of all unit elements.

To obtain $\epsilon'\epsilon$ minimum under this condition, let

$$\psi = \epsilon'\epsilon + \lambda 1'D_b 1, \tag{3}$$

where λ is a Lagrangian multiplier. From (2)

$$\psi = (B'Z'ZB - B'Z'W - W'ZB + W'W) + \lambda 1'D_b 1. \tag{4}$$

From the definitions above

$$Z'Z = \rho, \tag{5}$$

$$Z'W = \rho_c, \tag{6}$$

$$W'W = 1. \tag{7}$$

Substituting (5), (6), and (7) in (4)

$$\psi = B'\rho B - B'\rho_c - \rho_c' B + 1 + \lambda 1'D_b 1. \tag{8}$$

In (1) it is shown that

$$\rho_c = \delta^{-1/2} r_c \tag{9}$$

and

$$\rho = \delta^{-1/2}(r - D_u + D_u D_a D_b^{-1})\delta^{-1/2}, \tag{10}$$

where we define $D_u = I - D_{rii}$, a diagonal matrix of test unreliability coefficients. Let

$$B = \delta^{1/2}\beta. \tag{11}$$

Substituting (9), (10), and (11) in (8)

$$\psi = \beta'(r - D_u + D_u D_a D_b^{-1})\beta - \beta' r_c - r_c'\beta + 1 + \lambda 1'D_b 1. \tag{12}$$

The unknowns on the right-hand side of (12) are β, D_b and λ.

Differentiating (12) with respect to β' and equating the resulting expression to zero to get an extremum,

$$\frac{\partial \psi}{\partial \beta'} = r_c - (r - D_u + D_u D_a D_b^{-1})\beta = 0,$$

or

$$\beta = (r - D_u + D_u D_a D_b^{-1})^{-1} r_c \,. \tag{13}$$

Differentiating ψ with respect to the scalars, b_i, $(i = 1, 2, \cdots, n)$ and equating the n resulting expressions to zero

$$\frac{\partial \psi}{\partial b_i} = \lambda - \beta_i^2 (u_i a_i)/b_i^2 = 0 \tag{14}$$

or

$$b_i = \beta_i (u_i a_i)^{1/2}/\lambda^{1/2}. \tag{15}$$

Summing these n equations

$$\sum b_i = \sum \beta_i (u_i a_i)^{1/2}/\lambda^{1/2}.$$

Thus in matrix notation we obtain

$$\lambda^{1/2} = 1'(D_u D_a)^{1/2}\beta / 1' D_b 1. \tag{16}$$

Substituting for $\lambda^{\frac{1}{2}}$ in (15) and collecting these n expressions as the diagonal matrix D_b we obtain

$$D_b = D_\beta (D_u D_a)^{1/2} \frac{1' D_b 1}{1'(D_u D_a)^{1/2}\beta}, \tag{17}$$

where D_β is a diagonal matrix with the β_i as diagonal elements.

In (1) it is shown that

$$\beta = \left(r - D_u + \frac{[(D_u D_a)^{1/2} 1][1'(D_u D_a)^{1/2}]}{1' D_b 1}\right)^{-1} r_c \,. \tag{18}$$

Using (18) in (17) we can therefore solve for D_b, the new test lengths. The new multiple correlation is given by

$$R_b^2 = \beta' r_c \,. \tag{19}$$

Next we extend the procedure to the case of differential prediction. Consider the following additions to the definitions given above. Let

N = the number of criteria,
W = an $(M \times N)$ matrix of criterion scores whose elements are deviate scores of the form $(w_{ij} - \bar{w}_j)/(\sqrt{M}\,\sigma_{w_j})$,
H = an $(M \times N^2)$ matrix consisting of difference vectors for all possible pairs of criterion vectors i and j, including $i = j$,
B = an $(n \times N^2)$ matrix of "least squares" regression vectors for estimating H from Z,
r_c = the $(n \times N)$ matrix of validity coefficients with the tests of original lengths,
ρ_c = the $(n \times N)$ matrix of validity coefficients with the tests of altered lengths.

From the differential prediction procedure (2) we have

$$\phi = 1'D_c1 - (1'C1/N), \tag{20}$$

the index of differential prediction efficiency, where

$$C = r_c'r^{-1}r_c \tag{21}$$

and D_c is a diagonal matrix of the diagonals of C.
Let

$$E = ZB - H \tag{22}$$

and

$$F_i = e_i1' - I \tag{23}$$

where e_i is a column vector of all zero elements except the ith, which is unity.
Let

$$G' = (F_1, F_2 \cdots, F_N). \tag{24}$$

Thus we have

$$H = WG' \tag{25}$$

and

$$E = ZB - WG'. \tag{26}$$

From (23), (24), and (26) postmultiplied by G and divided by $2N$ we obtain

$$\epsilon = (EG/2N) = (ZBG/2N) - W[I - (11'/N)] \tag{27}$$

since $G'G = 2N[I - (11'/N)]$. It can be shown that the trace of $\epsilon'\epsilon$ is equivalent to the trace of $E'E$. Let

$$BG/2N = J. \tag{28}$$

Let

$$W[I - (11'/N)] = t. \tag{29}$$

Then

$$\epsilon = ZJ - t. \tag{30}$$

We wish to minimize the trace of $\epsilon'\epsilon$ with the constraining condition

$$1'D_b1 = T.$$

Let

$$\psi = \operatorname{tr} \epsilon'\epsilon + \lambda 1'D_b1. \tag{31}$$

Let

$$\gamma_c = \rho_c[I - (11'/N)]. \tag{32}$$

Substituting (5), (6), (9), (10), (28), (29), (30), and (32) in (31) we obtain

$$\psi = \operatorname{tr} [J'\delta^{-1/2}(r - D_u + D_uD_aD_b^{-1})\delta^{-1/2}J - J'\gamma_c - \gamma_c'J + t't] + \lambda 1'D_b1. \tag{33}$$

Let

$$\delta^{-1/2}J = L. \tag{34}$$

Let

$$\alpha_c = r_c[I - (11'/N)] = \delta^{1/2}\gamma_c \,. \tag{35}$$

Let

$$R = r - D_u \,. \tag{36}$$

Let

$$\Delta = D_u D_a \,. \tag{37}$$

Substituting (34), (35), (36), and (37) in (33) we obtain

$$\psi = \operatorname{tr}\,[L'(R + \Delta D_b^{-1})L - L'\alpha_c - \alpha_c' L + t't] + \lambda 1' D_b 1. \tag{38}$$

Differentiating (38) with respect to row vectors of L' and equating the results to 0 we obtain

$$\frac{\partial\psi}{\partial L'} = \alpha_c - (R + \Delta D_b^{-1})L = 0$$

or

$$\alpha_c = (R + \Delta D_b^{-1})L. \tag{39}$$

Differentiating (37) with respect to D_b and equating the results to 0 we obtain

$$\frac{\partial\psi}{\partial D_b} = \lambda I - D_{LL'}\Delta D_b^{-2} = 0, \tag{40}$$

where $D_{LL'}$ is a diagonal matrix whose non-zero elements are the diagonal elements of LL'. Hence

$$D_b = (D_{LL'}\Delta)^{1/2}/\lambda^{1/2}. \tag{41}$$

It can be shown that

$$\lambda^{1/2} = \frac{1'(D_{LL'}\Delta)^{1/2}1}{1'D_b 1} = \frac{1'(D_{LL'}\Delta)^{1/2}1}{T}. \tag{42}$$

Substituting (42) in (41),

$$D_b = (D_{LL'}\Delta)^{1/2}\,\frac{T}{1'(D_{LL'}\Delta)^{1/2}1}. \tag{43}$$

From (39)

$$L = (R + \Delta D_b^{-1})^{-1}\alpha_c \,. \tag{44}$$

Let

$$L_i = (R + \Delta D_{bi}^{-1})^{-1}\alpha_c \,, \tag{45}$$

where

$$D_{b_1} = \frac{T}{1'D_a 1}\,D_a \tag{46}$$

and

$$D_{b_{i+1}} = \frac{(D_{L_iL_i'}\Delta)^{1/2}T}{1'(D_{L_iL_i'}\Delta)^{1/2}1}. \tag{47}$$

Using (45), (46), and (47) as a basis of successive approximations to L_i and D_{b_i}, continue until D_{b_i} stabilizes satisfactorily.

Now the regression vectors for the optimal test lengths will be given by

$$Y = \rho^{-1}\rho_c \,. \tag{48}$$

From (9), (10), (36), (37), and (48)

$$Y = \delta^{1/2}(R + \Delta D_b^{-1})^{-1}r_c \,. \tag{49}$$

But from (35) and (39)

$$L = (R + \Delta D_b^{-1})^{-1}r_c[I - (11'/N)]. \tag{50}$$

From (50)

$$\delta^{1/2}L = \delta^{1/2}(R + \Delta D_b^{-1})^{-1}r_c - \delta^{1/2}(R + \Delta D_b^{-1})^{-1}r_c(11'/N). \tag{51}$$

From (49) and (51)

$$Y = \delta^{1/2}[L + (R + \Delta D_b^{-1})^{-1}r_c(11'/N)]. \tag{52}$$

Furthermore the index of differential prediction efficiency "ϕ" as defined in (2) can be shown to be

$$\phi = \operatorname{tr} L'\alpha_c \,. \tag{53}$$

The procedures outlined in Section II may be related to the above mathematical development as follows:

Table 1 is given by (36).

Step 1 is based on (35).

Step 2 is based on (37).

Step 3 is based on (46).

Step 4 consists of calculating $D_{b_1}^{-1}$ from D_{b_1}.

Step 5 consists of calculating $\Delta D_{b_1}^{-1}$.

Step 6 consists of calculating the parenthetical term on the right side of (45) for $i = 1$.

Step 7 consists of calculating L_1 from (45).

Step 8 consists of calculating D_{b_2} from (47).

Step 9 consists of calculating an L_2 matrix from (45).

Step 10 uses equation (47) to calculate D_{b_2}.

In general steps 6 and 7 are repeated for successive values of i in (45) and step 8 is repeated for successive values of i in (47).

Step 11 uses (53) to get successive values of ϕ.

REFERENCES

1. Horst, Paul. Determination of optimal test length to maximize the multiple correlation. *Psychometrika*, 1949, 14, 79–88.
2. Horst, Paul. A technique for the development of a differential prediction battery. *Psychological Monographs*, 1954, 68, No. 9 (Whole No. 380).
3. Horst, Paul. Servant of the human sciences, Chapter 21, Section 7. Unpublished manuscript. Division of Counseling and Testing Services, University of Washington.

Manuscript received 12/6/54
Revised manuscript received 1/28/55

PSYCHOMETRIKA—VOL. 21, NO. 1
MARCH, 1956

THE REGRESSION OF GAINS UPON INITIAL SCORES

R. F. GARSIDE
UNIVERSITY OF DURHAM, ENGLAND

A method of estimating the regression of gains upon initial scores is suggested and compared with two other methods which have been used in recent investigations.

I. *Introduction*

Thorndike (12) pointed out in 1924 that the correlation between obtained gains and initial scores tended to be negative when the correlation between true gains and true initial scores was not negative. In the same year Thomson (10) showed that this tendency was due to the same errors of measurement occurring in both gains and initial scores with different sign. He derived a formula for calculating the correlation between true gains and true initial scores. Thomson (11) and Zieve (16) have given alternative forms of Thomson's original formula. They did not, however, consider curvilinear regression.

The question of the regression of gains upon initial scores has arisen again recently in several papers concerned with the effect of practice and coaching on intelligence test scores (7, 13, and 14). It therefore seems appropriate to consider the measurement of this regression, both linear and curvilinear.

II. *Linear Regression of Gains upon Initial Scores*

If the same test is given twice, or two parallel tests as defined by Gulliksen (3, p. 11) are given to the same group of testees, then

$$Y = X + G, \tag{1}$$

where X = true initial score, Y = true final score and G = true gain. Now

$$r_{GX} = \frac{\sigma_Y r_{XY} - \sigma_X}{\sqrt{\sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y r_{XY}}} \tag{2}$$

and, therefore,

$$b_{GX} = b_{YX} - 1. \tag{3}$$

It is clear from (2) that if $r_{XY} = 1$, then $r_{GX} = 1$. But if there is any variation in true gain between individuals within arrays of true initial score, then neither r_{XY} nor r_{GX} will equal unity. If, and only if, $r_{XY} = r_{GX} = 1$, then

$$b_{GX} = (\sigma_Y/\sigma_X) - 1. \tag{4}$$

In general, however, b_{GX} increases with r_{XY} as well as with the ratio of σ_Y to σ_X. Thus any estimate of the regression of gains upon initial scores should increase not only with the ratio of the standard deviations of final and initial scores but also with the correlation between these scores.

In practice, what we want to do is to measure the true gains at different levels of initial ability as indicated by initial test scores. That is, we want to be able to ascertain the regression of true gains upon obtained initial scores, whether the regression be linear or curvilinear. We shall deal with linear regression first. Now

$$r_{Gx} = \frac{\sigma_Y r_{xY} - \sigma_X r_{xX}}{\sigma_G}, \tag{5}$$

where x = obtained initial score, and, therefore,

$$b_{Gx} = \frac{\sigma_Y r_{xY} - \sigma_X r_{xX}}{\sigma_x}. \tag{6}$$

Assuming errors of measurement to be independent of initial scores, and applying the usual correction for attenuation, (6) becomes

$$b_{Gx} = \frac{\sigma_y r_{xy} - \sigma_x r_{xx'}}{\sigma_x}, \tag{7}$$

where y = obtained final score and $r_{xx'}$ = reliability of initial score. Thus,

$$b_{Gx} = b_{yx} - r_{xx'}. \tag{8}$$

The value of b_{Gx} obtained from (8) increases with r_{xy} and with the ratio of σ_y to σ_x. Errors of measurement do not tend to make b_{Gx} negative; they do not appear in G.

It should be noted that $r_{xx'}$ may not be equal to r_{xy}; if $r_{XY} < 1$, then

$$r_{xy}^2 < r_{xx'} r_{yy'}, \tag{9}$$

where $r_{yy'}$ = reliability of final score. Thus it is essential to obtain an independent measure of $r_{xx'}$. One of the internal consistency methods must be used for this purpose. If the split-half method is used, then the formula for the variance of b_{Gx}, derived below, is applicable.

The first half of the initial test must not be correlated with the second half. Odd items should be correlated with even items, so that individual variations in gain within the initial test itself will not reduce the correlation between its two halves. The obtained correlation must, of course, be corrected to give $r_{xx'}$. If the variance of b_{Gx} is to be calculated by the method derived below, the Spearman-Brown formula [Kelley (4, p. 406)] should be used in the form

$$r_{xx'} = \frac{4C_{xx'}}{V_x}, \tag{10}$$

where $C_{zz'}$ = covariance of z and z', the two halves of the initial test, and V_z = variance of the initial test.

It should be borne in mind that the split-half method only gives accurate results when the two halves of the test measure the same factor or factors, and when the test is a power test rather than a speed test. Cronbach and Warrington (2) have recently discussed the estimation of the reliability of time-limit tests.

III. *Variance of Regression Coefficient*

To test the significance of the linear regression of true gains upon initial scores, we require the variance of b_{Gz}. If the split-half method has been used to estimate $r_{zz'}$, then

$$V_{bGz} = V_{b_{yz}} + V_{4C_{zz'}/V_z} - 2C_{b_{yz}(4C_{zz'}/V_z)} \cdot \tag{11}$$

The variance of $4C_{zz'}/V_z$ can be found by Pearson's method (6, p. 492). Ignoring cubic terms, then it follows directly from Pearson's equation (*iii*) that

$$V_{4C_{zz'}/V_z} = \frac{(4C_{zz'})^2}{V_z^2}\left[\frac{V_{4C_{zz'}}}{(4C_{zz'})^2} + \frac{V_{V_z}}{V_z^2} - \frac{2C_{V_z 4C_{zz'}}}{4V_z C_{zz'}}\right]. \tag{12}$$

But

$$V_{4C_{zz'}} = 16V_{C_{zz'}} \tag{13}$$

and

$$\begin{aligned} C_{V_z 4C_{zz'}} &= 4C_{V_{[z+z']}C_{zz'}} = 4C_{[V_z+V_{z'}+2C_{zz'}]C_{zz'}} \\ &= 4C_{V_z C_{zz'}} + 4C_{V_{z'}C_{zz'}} + 8V_{C_{zz'}} \cdot \end{aligned} \tag{14}$$

According to Wishart (15, p. 44), as quoted by Kelley (4, p. 555), for a normal bivariate population,

$$N^2 V_{C_{12}} = (N-1)(\bar{V}_1\bar{V}_2 + \bar{C}_{12}^2), \tag{15}$$

where $\bar{V}$ and $\bar{C}$ indicate population variance and covariance, respectively, and

$$N^2 C_{V_1 C_{12}} = 2(N-1)\bar{V}_1\bar{C}_{12} \cdot \tag{16}$$

If N is large, we may substitute sample values for the parameters in (15) and (16) and write (13) and (14) as follows:

$$V_{4C_{zz'}} = \frac{16(V_z V_{z'} + C_{zz'}^2)}{N} \tag{17}$$

and

$$\begin{aligned} C_{V_z 4C_{zz'}} &= (8V_z C_{zz'} + 8V_{z'}C_{zz'} + 8V_z V_{z'} + 8C_{zz'}^2)/N \\ &= [8(V_z + C_{zz'})(V_{z'} + C_{zz'})]/N. \end{aligned} \tag{18}$$

Similarly, it follows directly from another of Wishart's equations (15) that

$$V_{V_x} = 2V_x^2/N. \tag{19}$$

Substituting (17), (18), and (19) in (12), simplifying, and collecting terms,

$$V_{4C_{xx'}/V_x} = \frac{16}{NV_x^3}[V_x(V_xV_{x'} + 3C_{xx'}^2) - 4C_{xx'}(V_x + C_{xx'})(V_{x'} + C_{xx'})]. \tag{20}$$

But

$$V_x = V_x + V_{x'} + 2C_{xx'}\,. \tag{21}$$

Therefore, substiuting (21) in (20), simplifying, and collecting terms,

$$\begin{aligned} V_{4C_{xx'}/V_x} &= \frac{16V_xV_{x'}}{NV_x^3}[V_x + V_{x'} - 2C_{xx'} - V_xr_{xx'}^2 - V_{x'}r_{xx'}^2 + 2C_{xx'}r_{xx'}^2] \\ &= \frac{16V_xV_{x'}}{NV_x^3}[V_x + V_{x'} - 2C_{xx'}][1 - r_{xx'}^2]. \end{aligned} \tag{22}$$

We may note, in passing, that if it is assumed that $V_x = V_{x'}$, then (22) simplifies to

$$V_{4C_{xx'}/V_x} = \frac{4(1 - r_{xx'})^2}{N(1 + r_{xx'})^2}\,, \tag{23}$$

which is the same result as obtained by Shen (8, p. 462), using quite a different method.

The covariance of b_{yx} and $4\,C_{xx'}/V_x$ can be derived by writing

$$b_{yx} = C_{xy}/V_x\,. \tag{24}$$

Then, by using Pearson's equation (v) (6, p.493),

$$C_{(C_{xy}/V_x)(4C_{xx'}/V_x)} = \frac{4C_{xy}C_{xx'}}{V_x^2}\left[\frac{C_{C_{xy}4C_{xx'}}}{4C_{xy}C_{xx'}} - \frac{C_{V_xC_{xy}}}{V_xC_{xy}} - \frac{C_{V_x4C_{xx'}}}{4V_xC_{xx'}} + \frac{V_{V_x}}{V_x^2}\right]. \tag{25}$$

Now

$$C_{C_{xy}4C_{xx'}} = 4C_{C_{[x+x']y}C_{xx'}} = 4C_{[C_{yx}+C_{yx'}]C_{xx'}} = 4C_{C_{yx}C_{xx'}} + 4C_{C_{yx'}C_{xx'}}\,. \tag{26}$$

But Wishart (15) has shown that, for a normal trivariate population,

$$N^2C_{C_{12}C_{13}} = (N - 1)(\tilde{V}_1\tilde{C}_{23} + \tilde{C}_{12}\tilde{C}_{13}). \tag{27}$$

Thus, assuming N is large, we may write

$$\begin{aligned} C_{C_{xy}4C_{xx'}} &= 4(V_xC_{yx'} + C_{yx}C_{xx'} + V_{x'}C_{yx} + C_{yx'}C_{xx'})/N \\ &= 4[C_{yx'}(V_x + C_{xx'}) + C_{yx}(V_{x'} + C_{xx'})]/N. \end{aligned} \tag{28}$$

Moreover, it follows from (16) that

$$C_{V_xC_{xy}} = 2V_xC_{xy}/N. \tag{29}$$

Therefore, substituting (18), (19), (28), and (29) in (25) and simplifying,

$$C_{(C_{zy}/V_z)(4C_{zz'}/V_z)} = \frac{4}{NV_z^3}[V_zC_{yz'}(V_z + C_{zz'}) + V_zC_{yz}(V_{z'} + C_{zz'}) - 2C_{zy}(V_z + C_{zz'})(V_{z'} + C_{zz'})] \tag{30}$$

$$= \frac{4}{NV_z^3}[V_zC_{zz'}(C_{yz'} + C_{yz}) + V_z(V_zC_{yz'} + V_{z'}C_{yz}) - 2C_{zy}(V_z + C_{zz'})(V_{z'} + C_{zz'})].$$

But

$$C_{yz'} + C_{yz} = C_{zy}\,. \tag{31}$$

Substituting (21) and (31) in (30), multiplying out, and collecting terms,

$$C_{(C_{zy}/V_z)(4C_{zz'}/V_z)} = \frac{4}{NV_z^3}[V_zC_{yz'}(V_z + 2C_{zz'}) + V_{z'}C_{yz}(V_{z'} + 2C_{zz'}) - C_{zy}(V_zC_{zz'} + V_{z'}C_{zz'} + V_zV_{z'})]$$

$$= \frac{4}{NV_z^3}[V_zC_{yz'}(V_z - V_{z'}) + V_{z'}C_{yz}(V_z - V_z) - C_{zy}(V_zC_{zz'} + V_{z'}C_{zz'} + V_zV_{z'})]$$

$$= \frac{4}{NV_z^3}[V_z(V_zC_{yz'} + V_{z'}C_{yz}) - V_zV_{z'}(C_{yz'} + C_{yz}) - C_{zy}(V_zC_{zz'} + V_{z'}C_{zz'} + V_zV_{z'})] \tag{32}$$

$$= \frac{4}{NV_z^2}[V_zC_{yz'} + V_{z'}C_{yz} - b_{yz}(V_zC_{zz'} + V_{z'}C_{zz'} + 2V_zV_{z'})]$$

$$= \frac{4V_zV_{z'}}{NV_z^2}[b_{yz'} + b_{yz} - b_{yz}(b_{zz'} + b_{z'z} + 2)].$$

Thus, substituting (22) and (32) and the usual formula for $V_{b_{yz}}$ in (11), we get

$$V_{b_{Gz}} = \frac{1}{NV_z}\left\{V_y[1 - r_{zy}^2] + \frac{16V_zV_{z'}}{V_z^2}[V_z + V_{z'} - 2C_{zz'}][1 - r_{zz'}^2] - \frac{8V_zV_{z'}}{V_z}[b_{yz'} + b_{yz} - b_{yz}(b_{zz'} + b_{z'z} + 2)]\right\}. \tag{33}$$

Equation (33) involves the assumptions that N is large and that distributions are normal.

It has been pointed out to me that, by a theorem of Cramér (1, p. 366), any function of moments such as b_{Gz}, when N is large, is approximately

normally distributed about the corresponding population parameter. Thus, if N is large, b_{Gx} may be regarded as being distributed normally about $\tilde{b}_{Gx}$ with a variance as given by equation (33).

IV. *Curvilinear Regression*

When the regression of true gains upon initial scores is not linear, then a polynomial

$$G = a + bx + cx^2 + \cdots + px^n \tag{34}$$

can be used. But before the weighting coefficients b, c, etc. can be calculated, we must ascertain the relationships between them and the coefficients b', c', etc. of the polynomial

$$y = a' + b'x + c'x^2 + \cdots + p'x^n. \tag{35}$$

These coefficients may be calculated from the available data using the ordinary multiple correlation technique, which Kelley (4, p. 445) refers to in this connection as parabolic regression.

It follows from (6), by substituting x^a for x, multiplying by V_{x^a} and correcting for attenuation, that

$$C_{Gx^a} = C_{x^a y} - C_{x'x^a}\,. \tag{36}$$

$C_{x'x^a}$ in (36) is the covariance between the scores of a hypothetical application of the initial test and the ath power of the scores of *another* application of the initial test, a being an integer.

If the two applications of the initial test are parallel, then their standard deviations are equal. Thus, following Mollenkopf (5), as quoted by Gulliksen (3, pp. 119 and 120), we may write

$$x' = r_{xx'}x + e', \tag{37}$$

where x and x' = deviation scores of the initial test and e' = error of estimate when x' is predicted from x. Now

$$NC_{x'x^a} = \Sigma x'x^a. \tag{38}$$

Therefore, substituting (37) in (38),

$$NC_{x'x^a} = r_{xx'}\Sigma x^{a+1} + \Sigma e'x^a. \tag{39}$$

Let us assume that

$$r_{e'x^a} = 0. \tag{40}$$

From the raw score correlation formula

$$r_{e'x^a} = (\Sigma e'x^a - N\overline{e'}\,\overline{x^a})/(N\sigma_{e'}\sigma_{x^a}). \tag{41}$$

Thus

$$\Sigma e'x^a = N\overline{e'}\,\overline{x^a}. \tag{42}$$

But

$$Ne' = \Sigma e' = 0. \tag{43}$$

Therefore

$$\Sigma e' x^c = 0. \tag{44}$$

Thus, substituting (44) in (39),

$$C_{x'x^c} = r_{xx'} \Sigma x^{c+1}/N. \tag{45}$$

Therefore

$$C_{x'x^c} = r_{xx'} D_{xx^c}, \tag{46}$$

where D_{xx^c} = covariance between x and x^c. Therefore, substituting (46) in (36),

$$C_{Gx^c} = C_{x^c y} - r_{xx'} D_{xx^c}. \tag{47}$$

In matrix notation,

$$b_{Gx} = C_{Gx} C_{xx}^{-1}. \tag{48}$$

Therefore, substituting (47) in (48),

$$b_{Gx} = (C_{xy} - r_{xx'} D_{xx}) C_{xx}^{-1}, \tag{49}$$

where D_{xx} = the row matrix $[V_x \; C_{xx^2} \; \cdots \; C_{xx^n}]$, and $r_{xx'}$ = a scalar. Thus

$$b_{Gx} = b_{yx} - r_{xx'} D_{xx} C_{xx}^{-1}. \tag{50}$$

But

$$D_{xx} C_{xx}^{-1} = [1 \; 0 \; 0 \; \cdots \; 0]. \tag{51}$$

Therefore, reverting to ordinary notation,

$$b_{Gx \cdot x^2 \cdots x^n} = b_{yx \cdot x^2 \cdots x^n} - r_{xx'} \tag{52}$$

and

$$b_{Gx^c \cdot xx^d \cdots x^n} = b_{yx^c \cdot xx^d \cdots x^n}, \tag{53}$$

where c, d, and n are different integers not equal to unity.

It is evident from (52) and (53) that, although the regression lines of G and y on x do not have the same slope, they are, nevertheless, the same shape. If, therefore, the regression of y on x is not linear, then the departure from linearity must be due to the curvilinear regression of G on x. Thus, the usual test of departure from linear regression [Snedecor (**9**, section 14.4)] of y on x provides a test of the significance of the curvilinear regression of G on x. If the regression of y on x is significantly non-linear, it may be found that a second-degree polynomial will fit the data satisfactorily. Whether the inclusion of x^3 and higher powers of x produces any significant improvement in the prediction of y, and therefore of G, can be tested by analysis of variance [Kelley (**4**, p.448)].

It may be argued against the method here proposed that gains should

not be forced on to a curve of certain shape. But if a polynomial of sufficiently high degree has been used so that the deviations of y, and therefore of G, from the resulting curve are not significant, then such deviations can be ignored. If, however, there are a priori reasons for taking such non-significant deviations into account, then a polynomial of higher degree can be fitted.

V. *Comparison with Other Methods*

Wiseman and Wrigley (14) used three methods of calculating differential gains but finally decided to base their conclusions on the method which they describe as follows:

> The "level" of ability was defined by taking the sum of the quotients for the initial and final tests. Thus, if x is the score on the initial test and y the score on the final test, $(x + y)$ was found for each child. The $x + y$ scores were now placed in rank order and divided into levels (e.g. 180–199, 200–219), and the average gain $(y - x)$ for each level calculated. By plotting these points graphically it can easily be seen whether there is any rise or fall in $(y - x)$ with increase in $(y + x)$, and by fitting a straight line, or curve, to the points, the gain at any particular ability level may be estimated. It is necessary, of course, to re-translate the $x + y$ score into an I.Q. score on the first test in order to make the results meaningful. One way of doing this would be to find the average value of x (I.Q. on initial test) for each class-interval of $x + y$. A preferable method is to find equivalent x values from the regression of x on $(x + y)$. This was done.

Now

$$r_{(y-x)(y+x)} = (V_y - V_x)/(\sigma_{(y-x)}\sigma_{(y+x)}), \tag{54}$$

and

$$r_{x(y+x)} = (\sigma_y r_{xy} + \sigma_x)/\sigma_{(y+x)}, \tag{55}$$

$$E_x = b_{x(y+x)}(y + x) + c; \tag{56}$$

where E_x = estimated value of x, and c = a constant.

Having made a correlation scatter of $(y - x)$ against $(y + x)$, the correlation $r_{(y-x)(y+x)}$ is not changed by the substitution of E_x for $(y + x)$, since the correlation scatter itself remains unchanged. Therefore we may write

$$r_{(y-x)E_x} = r_{(y-x)(y+x)}. \tag{57}$$

Now

$$\sigma_{E_x} = \sigma_x r_{x(y+x)}. \tag{58}$$

Thus, substituting (55) in (58),

$$\sigma_{E_x} = \sigma_x(\sigma_y r_{xy} + \sigma_x)/\sigma_{(y+x)}. \tag{59}$$

Now

$$b_{(y-x)E_x} = r_{(y-x)E_x}\sigma_{(y-x)}/\sigma_{E_x}. \tag{60}$$

Therefore, substituting (54), (57), and (59) in (60),

$$b_{(y-x)E_x} = \frac{V_y - V_x}{\sigma_x(\sigma_y r_{xy} + \sigma_x)}. \tag{61}$$

Therefore

$$b_{(y-x)E_s} = \frac{(V_y/V_x) - 1}{b_{yx} + 1}. \tag{62}$$

The coefficient $b_{(y-x)E_s}$ is an estimate of the linear regression coefficient of gains upon initial score obtained by Wiseman and Wrigley. Errors of measurement do not tend to make this coefficient negative and it *increases* as r_{xy} *decreases*.

Peel (7) describes the percentile method as follows: "After calculating the scores at given percentiles in each of the three distributions, the differences between these 'equivalent levels' are then taken as the practice effects." These practice effects at each given percentile are then compared with corresponding initial scores. Assuming that the distributions are normal, the percentile method will give results as given in Table 1.

TABLE 1

Theoretical Differential Gains

Percentile	Initial Test (x)	Final Test (y)	Gain (g = y-x)
98th	$\bar{x} + 2\sigma_x$	$\bar{y} + 2\sigma_y$	$(\bar{y}-\bar{x}) + 2(\sigma_y-\sigma_x)$
84th	$\bar{x} + \sigma_x$	$\bar{y} + \sigma_y$	$(\bar{y}-\bar{x}) + (\sigma_y-\sigma_x)$
50th	$\bar{x}$	$\bar{y}$	$(\bar{y}-\bar{x})$
16th	$\bar{x} - \sigma_x$	$\bar{y} - \sigma_y$	$(\bar{y}-\bar{x}) - (\sigma_y-\sigma_x)$
2nd	$\bar{x} - 2\sigma_x$	$\bar{y} - 2\sigma_y$	$(\bar{y}-\bar{x}) - 2(\sigma_y-\sigma_x)$

We can therefore write

$$g = \frac{\sigma_y - \sigma_x}{\sigma_x}(x - \bar{x}) + (\bar{y} - \bar{x}) \tag{63}$$

and, therefore, assuming distributions are normal,

$$b_{gx} = (\sigma_y/\sigma_x) - 1. \tag{64}$$

It is evident that errors of measurement will not tend to make this regression coefficient negative. Moreover, gains calculated by the percentile method are independent of r_{xy} but increase as the ratio of σ_y to σ_x increases.

If it is assumed that $r_{xy} = 1$, then equation (62) becomes

$$b_{(y-x)E_x} = \frac{(\sigma_y/\sigma_x)^2 - 1}{(\sigma_y/\sigma_x) + 1}$$
$$= (\sigma_y/\sigma_x) - 1$$
$$= b_{gx} . \tag{65}$$

Thus if r_{xy} is equated to unity the Wiseman-Wrigley and the percentile methods give identical results.

The estimates of linear regression of gains upon initial scores obtained by each of the three methods here considered increase as the ratio of σ_y to σ_x increases. Using the proposed method, the regression of gains also increases as the correlation between x and y increases. But the percentile method is independent of this correlation and in the Wiseman-Wrigley method the gain regression increases as this correlation decreases. The proposed method would seem, therefore, to be superior.

VI. *Example*

Suppose the following data are obtained: $\sigma_y = 12$, $\sigma_x = 10$, $r_{xy} = .80$, $C_{xx'} = 23$ (i.e., $r_{xx'} \doteq .85$), $N = 400$ and there is no evidence of curvilinear regression. Then, by (10), $r_{xx'} = (4 \times 23)/100 = .92$ and, by (8), $b_{Gx} = [(12 \times .8)/10] - .92 = .04$. According to the Wiseman-Wrigley method, using (62), the corresponding coefficient is .22 and the percentile method gives a coefficient of .20 [equation (64)].

If r_{xy} is .90, instead of .80, then $b_{Gx} = .16$. The Wiseman-Wrigley coefficient = .21 and the percentile coefficient remains unchanged at .20.

When r_{xy} is high, as will be the case if reliable tests are used, the Wiseman-Wrigley method and the percentile method give similar results. These methods, however, may give results dissimilar to those obtained by the method here proposed.

REFERENCES

1. Cramér, H. Mathematical methods of statistics. Princeton: Princeton Univ. Press, 1951.
2. Cronbach, L. J. and Warrington, W. G. Time limit tests: estimating their reliability and degree of speeding. *Psychometrika*, 1951, **16**, 167–188.
3. Gulliksen, H. Theory of mental tests. New York: Wiley, 1950.
4. Kelley, T. L. Fundamentals of statistics. Cambridge: Harvard Univ. Press, 1947.
5. Mollenkopf, W. G. Variation of the standard error of measurement. *Psychometrika*, 1949, **14**, 189–229.
6. Pearson, K. On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc. roy. Soc.*, 1897, **60**, 489–498.
7. Peel, E. A. Practice effects between three consecutive tests of intelligence. *Brit. J. educ. Psychol.*, 1952, **22**, 196–199.
8. Shen, E. The standard error of certain estimated coefficients of correlation. *J. educ. Psychol.*, 1924, **15**, 462–465.

9. Snedecor, G. W. Statistical methods (4th Ed.). Ames, Iowa: Iowa State College Press, 1946.
10. Thomson, G. H. A formula to correct for the effect of errors of measurement on the correlation of initial values with gains. *J. exp. Psychol.*, 1924, **7**, 321–324.
11. Thomson, G. H. An alternative formula for the true correlation of initial values with gains. *J. exp. Psychol.*, 1925, **8**, 323–324.
12. Thorndike, E. L. The influence of the chance imperfections of measures upon the relation of initial score to gain or loss. *J. exp. Psychol.*, 1924, **7**, 225–232.
13. Vernon, P. E. Practice and coaching effects in intelligence tests. *Educ. Forum*, 1954, **18**, 269–280.
14. Wiseman, S. and Wrigley, J. The comparative effects of coaching and practice on the results of verbal intelligence tests. *Brit. J. Psychol.*, 1953, **44**, 83–94.
15. Wishart, J. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 1928, **20A**, 32–52.
16. Zieve, L. Note on the correlation of initial scores with gains. *J. educ. Psychol.*, 1940, **31**, 391–394.

Manuscript received 8/6/54
Revised manuscript received 2/21/55

PSYCHOMETRIKA—VOL. 21, NO. 1
MARCH, 1956

A METHOD OF SCALOGRAM ANALYSIS USING SUMMARY STATISTICS*

BERT F. GREEN

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

A method of Guttman scalogram analysis is presented that does not involve sorting and rearranging the entries in the item response matrix. The method requires dichotomous items. Formulas are presented for estimating the reproducibility of the scale and estimating the expected value of the chance reproducibility. An index of consistency is suggested for evaluating the reproducibility. An illustrative example is presented in detail. The logical basis of the method is discussed. Finally, several methods are suggested for dealing with non-dichotomous items.

Guttman's scaling method, known as scalogram analysis (4), has become popular among social scientists. However, current techniques for scalogram analysis are cumbersome. They all deal directly with the raw data in the form of an item response matrix that has a row for each respondent and a column for each item response category. An entry in the matrix indicates whether a particular respondent gave a particular item response. Various procedures have been described for rearranging the rows and columns of the item response matrix, as well as for combining response categories, so that a response "parallelogram" is achieved with few deviations. Suchman (12) described the scalogram board procedure in which the response matrix is represented by buckshot placed in small indentations in a set of removable slats. The sorting is accomplished by interchanging these slats in their frame. Methods for tabulating the response matrix on IBM equipment have been described by Noland (10), Ford (2), and Kahn and Bodine (6). Paper and pencil methods have been described by Guttman (3) and Marder (8).

These techniques are not automatic, but require keen judgment concerning the kind of sorting likely to pay off. Furthermore, the techniques are cumbersome since each attempts to evaluate the complete raw data matrix without the aid of summary statistics. For large numbers of respondents, the task is overwhelming. Moreover, it is difficult to deal with more than 10-20 items with these procedures. [A method, called H-technique, for combining items before making the scalogram analysis has been reported by Stouffer, Borgatta, Hays, and Henry (11)].

*Lois K. Anderson assisted the author materially in the many computations required for this paper. The research reported in this paper was supported in part by the Department of Economics and Social Sciences at M.I.T. and in part, jointly, by the Army, Navy and Air Force under contract with the Massachusetts Institute of Technology.

The purpose of this paper is to present a relatively simple method of scalogram analysis in which summary statistics are used to compute a close approximation to the scale *rep*. (In this paper "rep" is used for "reproducibility.") In this method, which requires dichotomous items, there is no limitation on the number of respondents; its application to large numbers of items is relatively easy. The method is particularly well-suited to punched-card techniques of processing data, since one must merely count the number of respondents who gave the positive response to each item, and the number of respondents giving certain specified combinations of responses. Obtaining these summary statistics is a simple, routine, completely objective matter.

In a sense, the method proposed here removes scalogram analysis from the list of subjective, slightly mystical techniques available only to experienced practitioners and places it on the list of objective methods available to any statistical clerk. The method also substantially reduces the time required for analysis. It gains these advantages at the expense of providing only an approximation to the "true" rep. Certain high-order scale errors are ignored. However the approximation appears to be a very close one.

The Method

All items must be dichotomous. In the mathematical notation we will let k be the number of items, N be the number of respondents, i be a subscript referring to item i (where the items are in any arbitrary order), and g be a subscript referring to item g in rank order.

Step 1. Designate the positive response to each item by referring to the item content. The positive response designations should be consistent with the investigator's hypothesis concerning the dimension being scaled.

Step 2. For each item tabulate n_i, the number of respondents who gave the positive response to the item.

Step 3. Arrange the items in rank order according to their popularities, (n_i/N), with the *least* popular item getting rank k, and the *most* popular item getting rank *1*. If there are any ties, adopt an arbitrary order.

Step 4. Tabulate $n_{g+1,\bar{g}}$ for $g = 1, 2, \cdots, k - 1$. This is the number of respondents who gave the positive response to item $g + 1$ and the negative response to item g. If it is easier to tabulate $n_{g+1,g}$ or $n_{\overline{g+1},g}$, then the following identities can be used:

$$n_{i\bar{j}} = n_i - n_{ij} ;$$

$$n_{\bar{i}j} = n_{i\bar{j}} + n_j - n_i .$$

Step 5. Use either of the following two methods for estimating the rep.

A. Tabulate $n_{g+2,g+1,\bar{g},\bar{g}-1}$ for $g = 2, 3, \cdots, k - 2$. This is the number of respondents who gave the positive response to item $g + 2$, *and* the positive response to item $g + 1$, *and* the negative response to item g, and the negative

response to item $g - 1$. Estimate the rep from the formula

$$\text{Rep}_A = 1 - \frac{1}{Nk}\sum_{g=1}^{k-1} n_{g+1,\bar{g}} - \frac{1}{Nk}\sum_{g=2}^{k-2} n_{g+2,g+1,\bar{g},\overline{g-1}} .$$

B. Tabulate $n_{g+2,\bar{g}}$ for $g = 1, 2, \cdots, k - 2$. This is the number of respondents who gave the positive response to item $g + 2$ *and* the negative response to item g. Estimate the rep from the formula

$$\text{Rep}_B = 1 - \frac{1}{Nk}\sum_{g=1}^{k-1} n_{g+1,\bar{g}} - \frac{1}{N^2k}\sum_{g=2}^{k-2} n_{g+2,\bar{g}} n_{g+1,\overline{g-1}} .$$

Rep_A and Rep_B should yield very similar estimates. The choice depends primarily on the relative ease of the alternative tabulations. Rep_A has the advantage that it is known to be an overestimate of the true (sample) rep.

The standard error of either Rep_A or Rep_B is approximately given by

$$\hat{\sigma}_{\text{Rep}} \approx \sqrt{\frac{(1 - \text{Rep})(\text{Rep})}{Nk}}.$$

Step 6. (*Optional*). Estimate the rep that would be expected by chance if the items had their observed popularities but were mutually independent. The rep of independent items is estimated by the formula

$$\text{Rep}_I = 1 - \frac{1}{N^2k}\sum_{g=1}^{k-1} n_{g+1} n_{\bar{g}} - \frac{1}{N^4k}\sum_{g=2}^{k-2} n_{g+2} n_{g+1} n_{\bar{g}} n_{\overline{g-1}} .$$

(Note that $n_{\bar{g}} = N - n_g$.)

Compute the *Index of Consistency*,

$$I = \frac{\text{Rep} - \text{Rep}_I}{1 - \text{Rep}_I},$$

where rep is either Rep_A or Rep_B. The index I will be unity if the items are perfectly scalable and has an expected value of zero when the items are independent. If the items show some negative correlation in the sample, I will be negative. If desired, label the set of items "scalable" if I is greater than .50.

Step 7. Give each respondent a scale score that is the number of items to which he gave the positive response.

Illustrative Example

A set of hypothetical data with $N = 20$, and $k = 6$ will be used as an example of the application of the method. The hypothetical data are shown in Table 1. The tabulations for Steps 2, 4, 5*A*, and 5*B* are also shown in Table 1. We have put the items and the respondents in rank order in Table 1 only to provide an easy comparison with the usual sorting techniques. In carrying out the tabulations of Steps 2, 4, and 5, it is not at all necessary that the

TABLE 1

Data and Tabulation for Illustrative Example
(+ = positive response; − = negative response)

Respondents	Items (i) 2	1	3	4	5	6	Scale Scores
	Rank Order (g) 6	5	4	3	2	1	
1	+	+	+	+	+	+	6
2	+	+	+	+	−	+	5
3	+	−	+	+	+	+	5
4	+	+	−	+	+	+	5
5	−	+	+	+	+	+	5
6	+	+	−	−	+	+	4
7	−	+	+	−	+	+	4
8	+	−	+	+	−	+	4
9	−	−	+	+	+	+	4
10	−	+	−	+	+	+	4
11	+	+	+	−	−	−	3
12	−	+	+	−	−	+	3
13	−	−	+	−	+	+	3
14	−	−	−	+	+	+	3
15	−	−	−	+	+	−	2
16	−	−	−	−	+	+	2
17	−	−	−	−	+	+	2
18	−	−	−	−	+	−	1
19	−	−	−	−	−	+	1
20	−	−	−	−	−	−	0
(Step 2) n_g	7	9	10	10	14	16	
(Step 6) $n_{\bar{g}}$	13	11	10	10	6	4	
(Step 4) $n_{g+1,\ \bar{g}}$	−	2	3	4	2	2	
(Step 5A) $n_{g+2,\ g+1,\ \bar{g},\ \overline{g-1}}$	−	−	1	2	0	−	
(Step 5B) $n_{g+2,\ \bar{g}}$	−	−	2	4	4	1	

raw data be arranged with either items or respondents in any particular order. It is quite possible to work directly with the individual response sheets, or with their punched-card equivalents.

From the formulas in Step 5, we compute

$$\mathrm{Rep}_A = 1 - \frac{1}{120}(2 + 2 + 4 + 3 + 2) - \frac{1}{120}(0 + 2 + 1) = .867$$

$$\mathrm{Rep}_B = 1 - \frac{1}{120}(2 + 2 + 4 + 3 + 2) - \frac{1}{2400}(1\cdot 4 + 4\cdot 4 + 4\cdot 2) = .880.$$

(The actual reproducibility is .858. The large discrepancy between this figure and the estimates is due to the small N in this example.) For Step 6

we compute

$$\mathrm{Rep}_r = 1 - \frac{1}{2400}(14\cdot 4 + 10\cdot 6 + 10\cdot 10 + 9\cdot 10 + 7\cdot 11)$$

$$- \frac{1}{960{,}000}(10\cdot 10\cdot 6\cdot 4 + 9\cdot 10\cdot 10\cdot 6 + 7\cdot 9\cdot 10\cdot 10) = .826$$

$$I_A = \frac{.867 - .826}{1 - .826} = .236.$$

Since I is less than .50, the set of items is not "scalable." The scale scores are shown in Table 1.

Justification of the Method

Ordering the items. The simplicity of the method of scalogram analysis presented in this paper is due largely to the use of popularity to rank the items. Guttman and his followers have used the order of items that yielded the highest rep. In a large majority of the cases, this order turns out to be the popularity order. In the other cases, the difference in the rep for the "best" (highest rep) order and for the popularity order is very small. Thus almost nothing will be lost and great simplification will be gained by using the popularity order. An arbitrary order may be adopted for tied items.

It is not surprising that the popularity order is usually the "best" order. In a perfect Guttman scale, the rank order of the items *must* correspond with the popularity order. For imperfect data, one would still expect the popularity order to be "best" if the scale errors are independent. Slight inversions might be expected if items had very similar popularities but the effect of these inversions would be small. Very peculiar error patterns would be required to make the popularity order markedly inferior to the "best" order.

Estimating Rep. The formulas for estimating the rep are based on an analysis of the scale errors in a pattern of item responses. In a perfect Guttman scale, the items can be arranged in a rank order so that a person who responds positively to (or endorses, or agrees with) any item also responds positively to all items of lower rank order. For a five-item scale, six ideal response patterns would be possible: [+++++], [−++++], [−−+++], [−−−++], [−−−−+], and [−−−−−]. In each ideal response pattern there is a dividing point, or cut, such that all item responses to the left of the cut are −, and all items responses to the right of the cut are +. The number of scale errors in any other response pattern is determined by placing a cut so that the number of +'s to the left of the cut and the number of −'s to the right of the cut are minimized. All such "misplaced" responses are errors. For example, the pattern [−−++−] would have its cut between items 4 and 3, (items are numbered in *decreasing* order from left to right, i.e.,

5, 4, 3, 2, 1) and one error would be counted. The pattern [−++−−] could have its cut between items 5 and 4 or at the right of item 1. In either case there would be two errors.

In order to find a rule for counting the errors in any particular response pattern we must consider subpatterns of responses. First, consider a pair of *adjacent* items with the response subpattern (+−); i.e., the response to the higher ranking item is + and the response to the lower ranking item is −. We would place the cut either to the right or to the left of this pair, and would have one error from the pair in either case. We would not place the cut between the two items, since this would yield two errors. Next, consider the reduced response pattern formed by deleting such a pair from a complete response pattern. Clearly the number of errors in the complete pattern is exactly one more than the number in the reduced pattern, for when we have determined the location of the cut in the reduced pattern, the two deleted items can and must be replaced together on either side of this cut. We may successively reduce a response pattern by eliminating pairs of adjacent items with (+−) subpatterns, until there are no remaining errors in the reduced pattern. The number of pairs eliminated is then the number of scale errors in the complete response pattern. For example, in the response pattern [−++−−] we first eliminate the pair (3, 2), (items are numbered 5, 4, 3, 2, 1) leaving (−+−); then we eliminate (4, 1), leaving (−). Hence there are two errors in the original pattern.

In practice the first step is simultaneously to eliminate from the complete response pattern *all* pairs of adjacent items with (+−) subpatterns. These are first-order errors. Next, we simultaneously eliminate all (+−) subpatterns from the reduced pattern; these are the second-order errors. We continue with third-, fourth-, and higher-order errors. Note that second-order errors are represented in the complete response pattern by a sub-pattern (++−−). The middle two items in this subpattern represent a first-order error, while the other two items represent the second-order error that appears when the first-order error is deleted. That is, second-order errors always straddle first-order errors. In the same way, third-order errors straddle second- and first-order errors, as in (+++−−−) or (++−+−−). In general, higher-order errors always straddle lower-order errors.

The formula for rep is

$$\text{Rep} = 1 - \frac{E}{Nk}, \tag{1}$$

where E is the total number of errors, N is the number of respondents, and k is the number of items. Then, we have

$$\text{Rep} = 1 - \frac{1}{Nk}\sum(+-) - \frac{1}{Nk}\sum(++--) - \text{terms of higher order.} \tag{2}$$

Now the sum of all adjacent item errors $(+-)$ is

$$n_{2\bar{1}} + n_{3\bar{2}} + \cdots + n_{k,\overline{k-1}} = \sum_{g=1}^{k-1} n_{g+1,\bar{g}} \,. \tag{3}$$

Similarly,

$$\sum (++--) = \sum_{g=2}^{k-2} n_{g+2,g+1,\bar{g},\overline{g-1}} \,. \tag{4}$$

Since the higher-order error patterns occur very infrequently, we will disregard all terms higher than second order. Substituting (3) and (4) in (2) we obtain the formula for Rep_A that we presented in Step 5*A*.

Instead of tabulating $\sum(++--)$, we may estimate this quantity. We shall assume that the two errors represented by the pattern $(++--)$ are independent. That is, the probability of a response pattern $(+\ 0\ -\ 0)$, where 0 symbolizes either response, is independent of the probability of a response pattern $(0\ +\ 0\ -)$. Then the product of these two probabilities is the probability of a response pattern $(++--)$. Under this assumption,

$$\frac{n_{g+2,g+1,\bar{g},\overline{g-1}}}{N} = \frac{n_{g+2,\bar{g}}}{N} \cdot \frac{n_{g+1,\overline{g-1}}}{N} \,. \tag{5}$$

Hence,

$$\sum_{g=2}^{k-2} n_{g+2,g+1,\bar{g},\overline{g-1}} = \frac{1}{N} \sum_{g=2}^{k-2} n_{g+2,\bar{g}} n_{g+1,\overline{g-1}} \,. \tag{6}$$

Substituting (3) in (2) and substituting (6) in (4) and the result in (2) we have the formula for Rep_B given in Step 5*B*.

A cruder approximation to rep can be obtained by ignoring the second-order error patterns, i.e., by omitting the last term from Rep_A or Rep_B. This approximation may be satisfactory in some cases, but it can be improved considerably by including the last term.

The formulas for Rep_A and Rep_B have been checked empirically using data published by Suchman (**12**, **13**). Table 2 gives the correct rep and our estimated Rep_A and Rep_B for ten scales of dichotomous items. The average discrepany is .002 for Rep_A and .003 for Rep_B. It should be pointed out that the actual reps are not necessarily those reported by Suchman since we have dichotomized all items and have not rejected any items from the scale. For larger numbers of items, the discrepancies may be slightly larger than those obtained here.

The errors in the proposed estimates of rep occur because terms of sixth and higher orders are neglected. It is possible to estimate the extent of the error by assuming that scale errors are independent. Both empirical and analytic results show that the error is about $(1 - \text{Rep})^3/\text{Rep}$.

The variance of Rep_A or Rep_B can be calculated. The calculations are straightforward but long, and lead to a complicated formula that will not

TABLE 2

Empirical Comparison of $\underline{\text{Rep}_A}$ and $\underline{\text{Rep}_B}$ with Correct Rep

Scalogram No. and Page Reference to Suchman (12, 13)	k	N	Correct Rep	Rep_A	Rep_B
0. p. 118	12	100	0.895	0.897	0.898
1. p. 124	9	100	0.923	0.924	0.925
2. p. 126	14	100	0.870	0.871	0.876
3. p. 130	9	100	0.890	0.888	0.895
4. p. 134	5	100	0.962	0.962	0.964
5. p. 136	7	100	0.970	0.970	0.970
6. p. 138	7	100	0.929	0.929	0.929
7. p. 140	10	100	0.913	0.917	0.918
8. p. 146	8	100	0.825	0.831	0.829
9. p. 148	6	100	0.943	0.945	0.946

be presented here. A satisfactory approximation, suggested by Guttman (4), is given by the simple formula presented in Step 5 above. The factor of k in the denominator leads to very small sampling variances which tend to give the investigator a false sense of security. It should be noted that errors of unreliability are usually much larger than sampling errors in this situation. Indeed, the errors in our approximations for rep are of the same order of magnitude as the sampling standard deviation.

Chance Rep. One of the major advantages of the method of scalogram analysis described here is the ease with which the chance rep, Rep_I, can be obtained. Rep_I is the rep that would be expected by chance, if the items were actually independent, and is a function of the item popularities, or "marginals." To obtain Rep_I, we note that if the items are independent, then

$$n_{g+1,\bar{g}} = n_{g+1}n_{\bar{g}}/N; \qquad n_{g+2,g+1,\bar{g},\overline{g-1}} = n_{g+2}n_{g+1}n_{\bar{g}}n_{\overline{g-1}}/N^3.$$

Substituting these values in the formula for Rep_A, we obtain the formula for Rep_I shown in Step 6 above.

The hypothesis that Rep_A (or Rep_B) is not significantly larger than Rep_I can be tested by using the variance estimate presented in Step 5. Caution is necessary when such a test yields borderline significance because of the many approximations involved. It should be noted that a significant rep does not necessarily indicate a homogeneous scale. For homogeneity the inter-correlations of the items should be fairly high, as well as significantly non-zero. It is possible to construct an index of homogeneity that will be zero when Rep_A (or Rep_B) = Rep_I, and will be unity when $\text{Rep}_A = 1$. The index

I, presented in Step 6, has these properties. It will be affected very little by changes in the number of items, or in the item popularities. This index is very similar to Loevinger's (7) index of homogeneity and to Menzel's (9) coefficient of scalability, and is proposed only because it is easier to compute in the present circumstance.

According to Guttman (4), a set of items should meet several criteria in order to be considered "scalable," or homogeneous. These criteria were apparently generated from anxiety about the chance rep [Festinger (1)]. With the exception of the requirement about a random pattern of errors, each criterion is related to the problem of avoiding spurious results and achieving homogeneity. It seems natural to replace them by a single criterion concerning I: I should be .50 or more for scalability. This criterion appears to give roughly comparable results to the many criteria used heretofore, and will be helpful to those who desire to create a dichotomy of scales *vs.* nonscales. To this author, it seems preferable to evaluate a scale in terms of an index of consistency that varies along a continuum rather than in terms of an arbitrary dichotomy.

Scoring. Any scoring method may be used with the present method since none of the previous steps depends on the scale scores. The concep of rep refers to the reproduction of an individual's item responses from a knowledge of his scale score. The measure of rep is a measure of the success of this reproduction when the scores are so assigned that the number of errors is minimized. It follows that the logically consistent method of scoring respondents is to compare their responses with the scale types. Each scale type or perfect response pattern is assigned a number, usually the number of positive responses in the response pattern. Then each individual's response pattern is compared with the scale types. An individual's score is the number assigned to the scale type that his responses match with the fewest deviations.

When the scale is perfectly reproducible, the respondent's score as determined by this scale-type-comparison process is equivalent to the number of positive responses that he made. Happily, it has been found [Suchman (12)] that the number of positive responses is very highly correlated with the scale-type-comparison score, when the scale is not perfectly reproducible. It appears that very little precision will be lost in practice by using the simple scoring method of counting positive responses. Although the issue is somewhat academic for small numbers of items, the savings of time and trouble are sizable for large numbers of items.

Non-dichotomous Items

In order to use the present method of analysis with items that have more than two response alternatives, the investigator must dichotomize the items. Several possibilities exist. He may use some logical a priori considerations, perhaps depending on the content of the alternative item responses.

The splits might be determined in part by the popularities of the alternatives. It would be possible to use some method such as that proposed by Jackson (5) for attempting to choose the dichotomies that maximize rep. For example, each respondent could be given a provisional scale score, and each possible dichotomy of each item could be correlated with this score; the dichotomies with highest correlations or covariances would be chosen.

An alternative procedure is to use *all possible dichotomies*. That is, an n-alternative item is replaced by $(n - 1)$ dichotomous items, or pseudo-items. For this procedure, it is necessary to adjust the value of k in the formulas. Wherever k appears as a factor, in conjunction with N, it remains the number of original non-dichotomous items, but in the limits of summation, k becomes the total number of pseudo-items. The pseudo-item trick will provide a crude approximation to the rep of a set of non-dichotomous items. This approximation is always slightly too low, but the value of I should be more accurate since the same bias exists in both Rep_A (or Rep_B) and Rep_I .

REFERENCES

1. Festinger, L. The treatment of qualitative data by "scale analysis." *Psychol. Bull.*, 1947, **44**, 146–161.
2. Ford, R. N. A rapid scoring procedure for scaling attitude questions. *Publ. Opin. Quart.*, 1950, **14**, 507–532.
3. Guttman, L. The Cornell technique for scale and intensity analysis. *Educ. psychol. Measmt.*, 1947, **7**, 247–279.
4. Guttman, L. The basis for scalogram analysis. In S. A. Stouffer *et al.*: Measurement and prediction. Princeton, N. J.: Princeton Univ. Press, 1950. Pp. 60–90.
5. Jackson, J. M. A simple and more rigorous technique for scale analysis. In: A manual of scale analysis, Part II. Montreal: McGill Univ., Mimeo, 1949.
6. Kahn, L. H. and Bodine, A. J. Guttman scale analysis by means of IBM equipment. *Educ. psychol. Measmt.*, 1951, **11**, 298–314.
7. Loevinger, Jane. A systematic approach to the construction and evaluation of tests of ability. *Psychol. Monogr.*, 1947, **61**, No. 4.
8. Marder, E. Linear segments: a technique for scalogram analysis. *Publ. Opin. Quart.*, 1952, **16**, 417–431.
9. Menzel, H. A new coefficient for scalogram analysis. *Publ. Opin. Quart.*, 1953, **17**, 268–280.
10. Noland, E. W. Worker attitude and industrial absenteeism: a statistical appraisal. *Amer. sociol. Rev.*, 1945, **10**, 503–510.
11. Stouffer, S. A., Borgatta, E. F., Hays, D. G., and Henry, A. F. A technique for improving cumulative scales. *Publ. Opin. Quart.*, 1952, **16**, 273–291.
12. Suchman, E. A. The scalogram board technique for scale analysis. In S. A. Stouffer *et al.*: Measurement and prediction. Princeton, N. J.: Princeton Univ. Press, 1950. Pp. 91–121.
13. Suchman, E. A. The utility of scalogram analysis. In S. A. Stouffer *et al.*: Measurement and prediction. Princeton, N. J.: Princeton Univ. Press, 1950. Pp. 122–171.

Manuscript received 11/8/54
Revised manuscript received 3/4/55

PSYCHOMETRIKA—VOL. 21, NO. 1
MARCH, 1956

NOTE ON CARROLL'S ANALYTIC SIMPLE STRUCTURE

HENRY F. KAISER
UNIVERSITY OF CALIFORNIA

A method for computing the transformation matrix for Carroll's analytic simple structure (1) is presented. The procedure involves successively finding the smallest latent root and associated vector of symmetric matrices.

A mathematical method of carrying out the iterations required in applying Carroll's analytic criterion for oblique simple structure (1) is developed. Since the trial and error procedure Carroll recommends occurs only in the computational stage of the solution, the general treatment of the problem will not be repeated. The notation used here is that of the original paper.

Carroll's criterion is that

$$f = \sum_{p=1}^{t-1} \sum_{q=p+1}^{t} \sum_{i=1}^{n} v_{ip}^2 v_{iq}^2 \tag{1}$$

$$= \sum_{p=1}^{t-1} \sum_{q=p+1}^{t} \sum_{i=1}^{n} \left(\sum_{m=1}^{s} a_{im}\lambda_{mp} \right)^2 \left(\sum_{m=1}^{s} a_{im}\lambda_{mq} \right)^2, \tag{2}$$

be a minimum, under the t conditions

$$g_p = \sum_{m=1}^{s} \lambda_{mp}^2 - 1 = 0. \tag{3}$$

He reaches a solution by systematically varying the elements of the transformation matrix Λ until the value of f attains a minimum. To this end, he is able to write

$$f_x = M_x' A M_R , \tag{4}$$

where

$$M_x' = [\lambda_{1x}^2 , \lambda_{2x}^2 , \cdots , \lambda_{sx}^2 , \lambda_{1x}\lambda_{2x} , \cdots , \lambda_{1x}\lambda_{sx} , \lambda_{2x}\lambda_{3x} , \cdots , \lambda_{2x}\lambda_{sx} , \cdots , \lambda_{(s-1)x}\lambda_{sx}] \tag{5}$$

is a row vector with $s(s + 1)/2$ variable elements, consisting of second-degree terms of the elements of column x of the transformation matrix Λ; where AM_R is a column vector whose elements are rather complicated functions of the elements of the arbitrary reference factor matrix F and the elements of the columns other than x in Λ; and where f_x is that part of

f which varies as a function of the elements of column x of Λ. Thus, in any one iteration f_x must be minimized with respect to λ_{mx}.

The $s(s+1)/2$ constant elements of AM_R are rewritten as

$$AM_R = \{b_{11}, b_{22}, \cdots, b_{ss}, 2b_{12}, \cdots, 2b_{1s}, 2b_{23}, \cdots, 2b_{2s}, \cdots, 2b_{(s-1)s}\}, \tag{6}$$

and the symmetric matrix B is defined as

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdot & \cdot & b_{1s} \\ b_{21} & b_{22} & \cdot & \cdot & b_{2s} \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ b_{s1} & b_{s2} & \cdot & \cdot & b_{ss} \end{bmatrix}, \qquad (b_{mk} = b_{km}). \tag{7}$$

[Note that the elements of the principal diagonal of B are the first s elements of AM_R; while the off-diagonal elements—with proper regard for subscripts—are one-half the value of the final $s(s-1)/2$ elements of AM_R.]

From (5) and (6), equation (4) becomes

$$f_x = \sum_{m=1}^{s} b_{mm}\lambda_{mx}^2 + 2\sum_{m=1}^{s-1}\sum_{k=m+1}^{s} b_{mk}\lambda_{mx}\lambda_{kx}. \tag{8}$$

But the λ_{mx} are constrained by (3), so that to minimize f_x let

$$\begin{aligned} 2\theta &= f_x - \mu g_x \\ &= \sum_{m=1}^{s} b_{mm}\lambda_{mx}^2 + 2\sum_{m=1}^{s-1}\sum_{k=m+1}^{s} b_{mk}\lambda_{mx}\lambda_{kx} - \mu\left(\sum_{m=1}^{s}\lambda_{mx}^2 - 1\right), \end{aligned} \tag{9}$$

where μ is a Lagrangian multiplier.

To find critical values of λ_{mx} set the partial derivative of (9), with respect to any one of the s variables λ_{mx}, equal to zero:

$$\partial\theta/\partial\lambda_{mx} = (b_{mm} - \mu)\lambda_{mx} + \sum_{\substack{k=1 \\ k\neq m}}^{s} b_{mk}\lambda_{kx} = 0. \tag{10}$$

If (10) is multiplied by λ_{mx}, and then summed over m,

$$\sum_{m=1}^{s} b_{mm}\lambda_{mx}^2 + 2\sum_{m=1}^{s-1}\sum_{k=m+1}^{s} b_{mk}\lambda_{mx}\lambda_{kx} = \mu\sum_{m=1}^{s}\lambda_{mx}^2. \tag{11}$$

Upon using (3) and (8),

$$\mu = f_x. \tag{12}$$

Thus writing out (10) explicitly for $m = 1, 2, \cdots, s$ gives

$$\begin{array}{lllll} (b_{11} - f_x)\lambda_{1x} + & b_{12}\lambda_{2x} & + \cdots + & b_{1s}\lambda_{sx} & = 0 \\ b_{21}\lambda_{1x} & + (b_{22} - f_x)\lambda_{2x} & + \cdots + & b_{2s}\lambda_{sx} & = 0 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ b_{s1}\lambda_{1x} & + \quad b_{s2}\lambda_{2x} & + \cdots + & (b_{ss} - f_x)\lambda_{sx} & = 0. \end{array} \tag{13}$$

Since these are homogeneous linear equations, the determinant of the coefficients of λ_{mx} must vanish:

$$\begin{vmatrix} (b_{11} - f_x) & b_{12} & \cdots & b_{1s} \\ b_{21} & (b_{22} - f_x) & \cdots & b_{2s} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ b_{s1} & b_{s2} & \cdots & (b_{ss} - f_x) \end{vmatrix} = 0. \tag{14}$$

Now (14) is obviously the characteristic equation of B, any latent root of which will make the determinant vanish. In order to minimize f_x, the smallest latent root is taken. The elements of the latent vector associated with this root are obviously the desired solutions for λ_{mx}.

For example, in Carroll's first iteration for Thurstone's box problem (1, p.34), B becomes

$$\begin{bmatrix} 3.931 & -.037 & .064 \\ -.037 & 1.705 & .144 \\ .064 & .144 & 1.263 \end{bmatrix}.$$

Employing Tintner's straightforward computational scheme for determining a smallest latent root and its associated latent vector (2, p.357-8),

$$\begin{aligned} \lambda_{1x} &= -.025 \quad (.000), \\ \lambda_{2x} &= -.287 \quad (-.292), \\ \lambda_{3x} &= .957 \quad (.956), \\ f_x &= 1.218 \quad (1.219). \end{aligned}$$

The results in parentheses are those obtained by Carroll by means of a trial and error procedure.

REFERENCES

1. Carroll, J. B. An analytic solution for approximating simple structure in factor analysis. *Psychometrika*, 1953, 18, 23–38.
2. Tintner, G. Econometrics. New York: Wiley, 1952.

Manuscript received 11/29/54
Revised manuscript received 3/21/55

BOOK REVIEWS

QUINN MCNEMAR. *Psychological Statistics, Second Edition.* New York, John Wiley and Sons, Inc., 1955, pp. vi + 408.

The second edition of McNemar's well-known text has the major virtues and faults of the first edition. Though the revisions are considerable, much of the original text remains unchanged, but practically all of the revisions represent improvements.

The primary virtue of this book—not shared, unfortunately, with some others in the same field—is its relatively complete freedom from major errors. The instructor who uses it will not have to spend much time explaining wherein and why the author's discussions are incorrect. Its primary fault is still its unevenness of exposition: careful, detailed, and lucid on many topics, condensed on several others to the point where students are unable to acquire any real understanding. Apparently the author could not quite bring himself to omit entirely a number of topics which he evidently did not feel he had space to develop fully. If the instructor is willing to skip over these topics, he will find McNemar's text excellent, but if he feels he must teach them thoroughly he will find it necessary to supplement the text frequently in his lectures and additional assignments.

The first notable improvement in the revised edition is a rearrangement of the materials of the earlier chapters, with considerable re-writing at various points. The discussion of the t-test now precedes that of correlation, so that all the materials on the sampling theory of one variable follow consecutively; the treatment of the binomial distribution has been expanded and used more fully as a starting point for the ideas of statistical inference. Though seven chapters on the one-variable case now replace six, the total number of pages has increased only from 108 to 114; the reviewer still doubts that this treatment is sufficient for students taking their first course in statistics.

The second major improvement is in the treatment of analysis of variance. A chapter (still too brief) on testing variances for homogeneity has been added, and the chapter on complex analysis of variance has been improved considerably.

A four-page chapter on distribution-free methods presents the sign test, the median test, Mood's test of C correlated sets, and the Mann-Whitney U-test. This chapter seems more like the result of an afterthought than a serious effort to discuss a now important area of modern statistical inference.

The nine-page chapter entitled "Distribution Curves" describes only the normal distribution and standard scores. The binomial distribution, and the normal approximation to it, are considered in the following chapter. The Poisson distribution is not discussed.

A few minor errors and inconsistencies still remain. Thus on page 100 the standard error of the mean of a sample from a finite population is given as $\sigma\sqrt{1 - n/N}/\sqrt{n}$ (where σ is the population standard deviation, N the population number, and n the sample number), instead of as $\sigma\sqrt{N - n}/\sqrt{n(N - 1)}$. Also on page 133 the standard error of estimate is given as $\sigma_{x \cdot y} = \sigma_x\sqrt{1 - r^2}$, omitting the factor $\sqrt{(N - 1)/(N - 2)}$. Similarly on pp. 138–9 the expression $1 - \sigma^2_{y \cdot x}/\sigma^2_y$ is defined as the proportion of the y-variance determined by x, again without the factor $(N - 1)/(N - 2)$, though the correct formula is implied in the case of multiple correlation on page 186. The so-called "shrunken" multiple correlation is in fact the square root of the coefficient of non-determination, and in the two-variable case this becomes $r'^2 = 1 - (1 - r^2)(N - 1)/(N - 2)$. McNemar calls the "shrunken" multiple correlation the unbiased estimate of the multiple correlation in the population; if this is so, r' as defined above is the unbiased estimate of ρ. The reviewer has never seen a proof that this *is* an unbiased estimate, and in fact it is probably not. For if we write it in the form, $r'^2 = 1 - s^2_{y \cdot x}/s^2_y$, $s^2_{y \cdot x}$ *is* an unbiased estimate of $\sigma^2_{y \cdot x}$, and s^2_y *is* an unbiased estimate of σ^2_y; the unbiased estimate of a ratio is seldom the quotient of unbiased estimates of its numerator and denominator, and the unbiased estimate of a square root is seldom the square root of

the unbiased estimate of the number. Aside from these and a few other liberties with such ratios as $N/(N-1)$ and $(N-1)/(N-2)$, however, there are few errors.

The formula for the large-sample standard error of $(r_{12} - r_{13})$ has been replaced in the revised edition by Hotelling's t-test for this difference.

To the present reviewer, the author's deliberate lack of emphasis on efficient computing methods appears to be a deficiency, but this is admittedly a matter of opinion wherein many other authors agree with McNemar.

In summary, those instructors who liked the first edition of McNemar's book will like the second edition better; those who did not like the first edition will probably have the same objections (though somewhat weakened) to the second. This reviewer, it may be added, is one of the former.

University of Tennessee

Edward E. Cureton

R. L. Thorndike and Elizabeth Hagen. *Measurement and Evaluation in Psychology and Education.* New York: Wiley, 1955, pp. vii + 575.

This book is an outgrowth of, and was developed for, a Teachers College course for teachers, administrators, guidance workers, and majors in several branches of psychology. "It undertakes to provide the foundations that these workers in different branches of education and psychology will need in order to use and interpret tests, move ahead into more specialized testing courses, and go ahead independently to study their own practical testing problem." (Preface, p. v). It seems admirably suited to these purposes; it should also be an excellent handbook for the school administrator who is not an expert on testing to peruse and to have available for reference.

Many value-judgments are made throughout the book, as is inevitable if it is to be practically serviceable, but I think most measurement people would agree with most of the judgments. Those to which I take most violent exception are noted below. The general effect, however, is a sound evaluation of tests in terms of what they can and cannot contribute to the making of decisions, principally in education.

In the first chapter, on historical and philosophical orientation, the authors depart from the usual nominal, ordinal, interval, and ratio categories of measurement, as proposed by Stevens, in favor of four other classes: either-or, qualitatively described degrees, rank in a group, and amount expressed in uniform established units. The first is identical with the Stevens nominal scale (one of the examples given refers to a man being either single, married, widowed, or divorced); the second and third are ordinal scales; and the fourth describes an interval scale, although the examples given—weight, height, and age—are ratio scales. The propaedeutic function of the book might have been better served by the use of standard terminology.

After a chapter cataloging the different measurement options with respect to what is measured and how it is to be done, there follow two on teachers' tests and on preparing objective tests. These chapters present a detailed plan for constructing a classroom test, built around a sample unit of instruction. Rules for item construction are given, together with examples of good and poor items.

The chapter on elementary statistical concepts reviews the usual topics of central tendency, variability, and relationship; in addition it points out the dependence of the usual interpretations of standard deviation upon the presence of a normal distribution. The discussion of the correlation coefficient and its interpretation is particularly good for the level of sophistication the authors have chosen.

The sixth chapter deals with test desiderata: validity, reliability, and practicality, and

culminates in a Schedule for Evaluating a Test. The discussion of reliability is especially good, but the treatment of validity differs from that of the recent APA Technical Recommendations, without any particular gain, it seems to me. Furthermore, the term "construct validity" is used with a meaning different from that of the APA committee.

The four common types of norms—age, grade, percentile, and standard-score—are next described and evaluated. In the discussion of standard scores the authors seem to go completely off the track, and are actually perpetuating the prevalent fallacy that there is some magic in the process of subtracting the mean and dividing by the standard deviation. To quote (p. 165): "Because the units of a score system based on percentiles are so clearly not equal, we are led to look for some other unit that does have the same meaning throughout the whole range of values. *Standard-score* scales have been developed to serve this purpose." An example is given in which Mary gets a score 1.50 above the mean on test A and Johnny, a member of the same class, gets the same standard score on test B. "Thus, we may say that Mary did as well on test A as Johnny did on test B . . ." (p. 166), To aid in interpreting the "degree of excellence represented by a standard score" the reader is referred to a table of percentile equivalents in a *normal* distribution. This puts the authors in the peculiar position of beginning the section on standard scores with a statement on the inadequacy of percentile scores, and then midway in the discussion using them to interpret the supposedly superior standard scores.

It is only several paragraphs beyond this that they take up the matter of normalized standard scores. In a summary, however, they say that standard scores are ". . . presumably equal units. The basic unit is the number of standard deviation units above or below the mean of the group" (p. 168).

It would seem to be too easy for the naive reader to get a false picture of the virtues of standard scores from this presentation. I should think it entirely within the grasp of the audience to whom the book is directed to understand the rationale for using a unit based on an assumed normal distribution of ability, and that non-normalized standard scores, being only linear transformations on raw scores, are no better than raw scores unless there is some kind of reference distribution or unless it is desired to compare performance across tests which have the same form of distribution.

This same chapter has an excellent discussion of profile interpretation, emphasizing the necessity for taking into account the reliabilities of difference scores.

Chapters 8 through 15 cover the topics which form the heart of the usual tests and measurements course: sources of information about tests, the various kinds of traits for which tests have been designed, and kinds of measuring instruments in current use. The general approach here has been to try to set forth principles governing various measurement techniques so as to give the reader a background for evaluation. Illustrative tests are discussed briefly, and findings of some of the research studies in relevant areas are summarized.

The last five chapters deal with planning a school testing program, marking and reporting, educational and vocational guidance, personnel selection, and diagnosis and therapy. In the chapter on Marking and Reporting, Thorndike and Hagen take the position that course marks can be only a relative appraisal, with respect to some reference group. They ignore the alternative of assigning marks based on the extent to which the students have achieved the operationally defined objectives of instruction. Sufficient progress has been made in this direction, certainly, to make it a functional alternative, and, for me at least, a preferable one.

There are four appendices, the first two of which are computational (square root and correlation coefficient). The third is a listing and evaluative description of some of the more widely used tests, and the fourth is a list of seven prominent test distributing agencies, with a description of the kinds of services they offer.

John E. Milholland

University of Michigan

Decision Processes, THRALL, R. M., COOMBS, C. H., AND DAVIS, R. L., Editors, John Wiley and Sons, New York, 1954, viii + 332, $5.00.

Decision Processes, while nominally a book, is in fact a one-issue journal consisting of nineteen mathematical and experimental papers on statistical decision theory, game theory, learning theory, and measurement theory (including utility measurement)—all parts of an area well described by the title. The work stems from an eight-week summer conference on "The Design of Experiments in Decision Processes" held at the RAND Corporation in 1952. On the grounds that such a book will not be definitive and that research activity in the area is lively, the editors felt that "an informal and relatively speedy method of printing" was justified. While agreeing with their conclusion, two questions can be raised: Did these considerations actually force the publisher to employ such an unattractive format? And do not these same reasons, plus the desirability of the lowest possible price for a volume soon to be antedated, suggest paper, not cloth, covers?

The volume begins with an introduction by R. L. Davis, which outlines the area, cites a bit of its history, and sketches the major focus and results of each paper. A clear notion of the relevance of this book to one's interests can be obtained by reading these eighteen pages. The next article, also introductory in nature, "Some Views on Mathematical Models and Measurement Theory" by C. H. Coombs, Howard Raiffa, and R. M. Thrall is divided into two parts. The first offers a highly idealized scheme of scientific research with particular emphasis on the role of mathematical models. The second part, on measurement models, is presented as an exemplification of the general scheme; it should serve as a handy reference of possible scales which, by being more complete, supplements Stevens' widely known classification. Definitions and social science illustrations are given of transitive relation, partial order, weak order, lattice, vector space, etc.; the interrelations among them are discussed and neatly summarized in a diagram.

The remaining articles are grouped in four sections: individual and social choice, learning theory, theory and applications of utility, and experimental studies. Since it is impossible to discuss them all in detail, attention will be restricted to those the reviewer found particularly satisfying or stimulating; as it happens all four divisions of the book are represented.

L. A. Goodman's paper "On Methods of Amalgamation," John Milnor's "Games Against Nature," and the "Note on Some Proposed Decision Criteria" by Roy Radner and Jacob Marschak are all concerned with decision criteria for the selection of a strategy in a game against nature. Goodman offers a new criterion which, simultaneously, generalizes those of LaPlace, Bayes, and Copeland. Radner and Marschak present an example which suggests that both the Hurwicz generalization of the Wald minimax criterion and the Savage minimax regret criterion may be inadequate, and, as we shall see, Milnor's work raises similar doubts. The Hurwicz criterion leads to a decision distinctly at variance with common sense, and the Savage criterion depends on irrelevant alternatives, in a sense analogous to Arrow's usage. Milnor's paper, the most interesting and elegant of the three, overlaps the others, covering the LaPlace, Wald, Hurwicz, and Savage criteria. Milnor lists eleven axioms a criterion might meet, and he shows which are met by the four criteria mentioned, and which characterize each of the four. It is striking that all but the LaPlace criterion fail to meet a Pareto condition on strategies (domination), and that the LaPlace criterion fails on another axiom, which, while not so basic, seems desirable. Furthermore, no criterion can meet all eleven axioms, so one is led to consider classes of criteria defined by subsets of axioms which seem intuitively necessary. Milnor selects five as essential and three others as desirable; he shows that the class so defined is non-empty. Finding a simple characterization of this class of criteria, or indeed of any member of the class, remains an unsolved problem.

The first paper of part II, "A Formal Structure for Multiple-Choice Situations"

by R. R. Bush, Frederick Mosteller, and G. L. Thompson, is a welcome concise statement of the mathematical structure of the Bush-Mosteller stochastic learning model. As is well known, the model can be stated in very general terms, but most of the results and applications assume linear operators. A major and controversial part of the paper is an attempt by means of the "combining of classes and condition," to give a more respectable basis for this assumption than the intriguing observation that it works. Roughly, this condition requires that the model yield the same results whether or not two alternatives with the same set of outcome probabilities are combined. At first glance this seems to have the same status and intuitive necessity as, say, the requirement that the laws of physics shall be independent of the position of the observer; to the extent it has this status and necessity it is exciting. Careful inquiry, however, suggests otherwise, for the probabilities relating outcomes to alternatives are under the arbitrary control of the experimenter; hence, the model must allow for *any* possible combining of classes. It appears to the reviewer that this is too demanding to be considered intuitively necessary, and thus is not really a justification for the linearity assumption. Still a persuasive justification is needed, for the linear model fits an impressive collection of data. An example of such data is presented in "Individual Behavior in Uncertain Situations: An Interpretation in Terms of Statistical Association Theory" by W. K. Estes.

Part III, on utility, includes two papers on the existence of utility functions; these papers are interesting but mathematically the most difficult in the book. The first, "Representation of a Preference Ordering by a Numerical Function" by Gerard Debreu, is concerned with topological conditions on a weakly-ordered set which are sufficient to insure the existence of a utility function. If certain sets are closed, he shows that either separability and connectedness or perfect separability are sufficient. No algebra of probability-combining is assumed as in the von Neumann and Morgenstern theory, but no unique results are obtained. In "Multidimensional Utilities" Melvin Hausner examines the effect of dropping the Archimedean axiom from the von Neumann and Morgenstern axioms. Let ApB denote a probability combination of A and B; the axiom requires that if A is preferred to B, and B to C, then ApC and B are indifferent for some p. The possible objection to the axiom is seen when one lets A = five cents, B = two cents, and C = death. Hausner obtains the elegant results that any non-Archimedean "mixture" space satisfying the other von Neumann and Morgenstern axioms can be imbedded in an ordered vector space, and that any ordered vector space is lexicographically ordered in some basis. Some interesting applications of this theory are suggested by R. M. Thrall in "Application of Multidimensional Utility Theory."

"Towards an Economic Theory of Organization and Information" by Jacob Marschak initiates a fascinating normative study of decision-making by communicating "teams," where teams are defined to be groups with identical individual and group utility functions. A team may collect data, transmit information over a communication network at some cost, and take actions based on a decision rule. Three classes of problems are considered for a team which completes all observation before making any decisions. 1) *Procedural*: given a network and cost of communication, to select the best rules for governing information transmission and actions. 2) *Network*: given rules and a cost function over networks, to select the best communication network. 3) *Constitutional*: to select the best procedural-network pair. Several simple special cases are solved, but as Davis notes (p. 13): "The relatively difficult manipulations required even for these simple cases show for one thing how desirable further development and simplification of the theory would be, while on the other hand they serve to emphasize how difficult would be any analysis at all without the machinery of this formalization."

In the final experimental section, two of the four papers deal with coalition formation in the game-theory sense; both emphasize that psychological rather than "objective"

utilities are necessary for a descriptive theory. In "Tendencies Toward Group Comparability in Competitive Bargaining," Paul Hoffman, Leon Festinger, and Douglas Lawrence employ Festinger's psychological theories of group behavior to predict that those who are perceived as superior in an ability relevant to the conflict of interest involved tend to be excluded from effective bargaining. The confirming experiment was based on a symmetric 3-person game. One player was always a stooge who, in one variation, appeared to be of similar intelligence to the subjects, but who, in the second variation, was evidently of superior ability. In the latter case he was excluded from coalitions more often than in the former, the degree increasing with the importance subjects placed on the game situation. These results strongly suggest that utility functions are subject to modification by psychological manipulations—an unfortunate complication. More directly related to game theory itself is the paper "Some Experimental n-Person Games" by G. Kalisch, J. W. Milnor, J. Nash, and E. D. Nering. Several n-person games ($n = 4, 5, 7$) were run in characteristic function form, i.e., payments were stated for each possible coalition. In each case subjects bargained for 10 minutes, and they reported their agreements to an umpire who enforced them. Considering the rationality assumptions of the theory, the time limit seems questionable. The principal results appear to be: contrary to theory, strategically equivalent games were treated differently; the Shapley value tended to be similar, though by no means identical, to the experimental payments; no satisfactory method was devised to check the von Neumann and Morgenstern theory of solutions. If the authors intended to show that objective payments rather than subjective utilities are sufficient for descriptive purposes, the first result is most disturbing. The failure of the subjects to respond to the objective situation is further confirmed by the authors' observation that the subjects tended to form coalitions having large payments without regard to benefits resulting from other *apparently* less impressive coalitions. While the prospects of positive findings are not great, the experiment probably should be replicated under more carefully controlled conditions and using many more subjects. At that time data could be collected from the subjects prior to each run as to their perceptions of relative coalition strength per coalition member. We do not expect these to be the same as the "rational" ordering derived from the objective characteristic function, but it might be possible to establish that their bargaining behavior is consistent with their orderings. Certainly these two experiments reinforce the contention of von Neumann and Morgenstern that an individual's utility function need not be simply related to any objective measure arising from the situation.

In summary, we may agree with the editors that the book is not definitive and yet recommend it as stimulating and useful for those working in the area. Anyone attracted by any one of the papers will surely be interested in several others, and he may very well have a passing curiosity about most of them.

Center for Advanced Study in the Behavioral Sciences

R. Duncan Luce

PSYCHOMETRIKA—VOL. 21, NO. 1
MARCH, 1956

MEMBERS OF THE PSYCHOMETRIC SOCIETY

Robert P. Abelson, Dept. of Psychology, Yale University, New Haven, Connecticut
Sidney Adams, 242 N. Granada Street, Arlington, Virginia
C. J. Adcock, Victoria University College, Wellington, New Zealand
Marvin Adelson, Hughes Aircraft Co., 5530 W. Century Boulevard, Los Angeles 45, California
Dorothy C. Adkins, Dept. of Psychology, University of North Carolina, Chapel Hill, North Carolina
Roger B. Allison, Jr., Research Dept., Educational Testing Service, Princeton, New Jersey
Earl A. Alluisi, P. O. Box 3109, University Station, Columbus 10, Ohio
Robert B. Ammons, Dept. of Psychology, University of Louisville, Louisville, Kentucky
Wade S. Amstutz, R. F. D. 1, Yellow Springs, Ohio
Anne Anastasi, Graduate School, Fordham University, New York 58, New York
Oliver F. Anderhalter, Bureau of Institutional Research, 15 N. Grand Boulevard, St. Louis 3, Missouri
Richard J. Anderson, 1203 12th Avenue, N. W., Gainesville, Florida
Scarvia B. Anderson, 337 Ottawa Street, Washington 21, D. C.
Gordon V. Anderson, Testing and Guidance Bureau, University of Texas, Austin 12, Texas
Nancy S. Anderson, U. S. Navy Electronics Laboratory, Code 2130, San Diego 52, California
T. W. Anderson, Dept. of Mathematical Statistics, Columbia University, New York 27, New York
T. G. Andrews, Dept. of Psychology, University of Maryland, College Park, Maryland
William H. Angoff, Educational Testing Service, Princeton, New Jersey
Valentine Appel, 3135 Decatur Avenue, New York 67, New York
E. James Archer, Dept. of Psychology, University of Wisconsin, Madison, Wisconsin
Thomas B. Armstrong, 39-35 51st Street, Apt. 5E, Woodside 77, Queens, New York
Philip Ash, 574 Mathews Street, Gary, Indiana
Melany E. Baehr, 109 Payne Avenue, Danville, Illinois
Andrew R. Baggaley, 600 W. Kilbourn Avenue, Milwaukee 3, Wisconsin
Donald E. Baier, Commonwealth Life Insurance Co., Louisville 2, Kentucky
Paul J. Barnes, 2436 Marion Avenue, New York 58, New York
A. S. Barr, University of Wisconsin, Madison 6, Wisconsin
Nancy Bayley, Psychology Laboratory, Institute of Mental Health, National Institute of Health, Bethesda 14, Maryland
David C. Beardslee, Psychological Laboratories, Wesleyan University, Middletown, Connecticut
Katharine Pease Beardsley, 3817 Van Ness Street, N. W., Washington 16, D. C.
Harold P. Bechtoldt, Dept. of Psychology, State University of Iowa, Iowa City, Iowa
Roger M. Bellows, Personnel Research Center, Inc., 5057 Woodward Avenue, Detroit 2, Michigan
A. W. Bendig, Dept. of Psychology, University of Pittsburgh, Pittsburgh 13, Pennsylvania
George K. Bennett, Psychological Corporation, 522 Fifth Avenue, New York 36, New York
Joseph F. Bennett, Farrar Road, S., Lincoln, Massachusetts
Purnell Benson, Drew University, Madison, New Jersey
Arnold Binder, Dept. of Psychology, Indiana University, Bloomington, Indiana
Joseph D. Birch, Dept. of Psychology, University of Michigan, Ann Arbor, Michigan

James E. Birren, National Institute of Mental Health, Bethesda 14, Maryland
Paul Blommers, State University of Iowa, Iowa City, Iowa
Frances Blumenthal, 6 Landscape Avenue, Yonkers 5, New York
R. Darrell Bock, 6109 S. Ellis Avenue, Chicago 37, Illinois
Nicholas A. Bond, Jr., 7111-A Alvern Street, Los Angeles 45, California
Edward S. Bordin, Bureau of Psychological Services, 1027 E. Huron Street, Ann Arbor, Michigan
Edgar F. Borgatta, 34-18 73rd Street, Jackson Heights, New York
Henry Borow, The General College, University of Minnesota, Minneapolis 14, Minnesota
Robert A. Bottenberg, 2666 W. Summit, San Antonio, Texas
Rev. W. A. Botzum, University of Portland, Portland 3, Oregon
Eugene A. Bouvier, Dept. of Psychology, University of Southern California, Los Angeles 7, California
James E. Boyce, 1905 Proctor, Flint 4, Michigan
Claude F. Bridges, 48 Post Avenue, New York 34, New York
Harold F. Bright, Technical Services, Human Resources Research Office, P. O. Box 3596, Washington 7, D. C.
Hubert E. Brogden, Personnel Research and Procedures Branch, AGO, War Dept., Pentagon, Washington 25, D. C.
Theodore H. Brown, Soldiers Field Road, Harvard Business School, Boston 63, Massachusetts
Virginia M. Brown, 1700 E. Tuft Avenue, Englewood, Colorado
William F. Brown, Training Methods Division, HumRRO, P. O. Box 3596, Washington 7, D. C.
Glenn L. Bryan, Dept. of Psychology, University of Southern California, 3516 University Avenue, Los Angeles 7, California
Joseph G. Bryan, 74 Springfield Street, Watertown 72, Massachusetts
Donald N. Buckner, 1728 Strand, Hermosa Beach, California
Hermann F. Buegel, University of North Dakota, Box 506, Grand Forks, North Dakota
Eugene I. Burdock, Carnegie Corporation of New York, 589 Fifth Avenue, New York 17, New York
George Graydon Burgess, 110A Gillmore Ave., B. Mitchell Vill., San Antonio, Texas
Laverne K. Burke, 133 Anacostia Road, S. E., Washington 19, D. C.
William W. Burke, School of Social Work, Washington University, St. Louis, Missouri
Oscar K. Buros, 220 Montgomery Street, Highland Park, New Jersey
Raymond H. Burros, Dept. of Psychology, University of Houston, Houston 4, Texas
Harold E. Burtt, Dept. of Psychology, Ohio State University, Columbus 10, Ohio
John Caffrey, Division of Guidance & Research, Office of the County Supt. of Schools, 808 N. Spring Street, Los Angeles, California.
Joel T. Campbell, 1556 Knuth Avenue, Euclid 32, Ohio
Frederick O. Carleton, 1812 C Street N. E., Washington, D. C.
John B. Carroll, 13 Kirkland Street, Cambridge 38, Massachusetts
Harold D. Carter, 10 E. Parnassus Street, Berkeley 8, California
Launor F. Carter, The Rand Corporation, 1700 Main Street, Santa Monica, California
Anthony Castlenovo, 11418 Bellflower Avenue, Personnel Research Institute, Western Reserve University, Cleveland 6, Ohio
Raymond B. Cattell, 106 S. Goodwin Street, Urbana, Illinois
Randall M. Chambers, Division of Behavior Studies, Jackson Memorial Laboratory, Box 847, Bar Harbor, Maine
Robert L. Chapman, Rand Corporation, 1500 Fourth Street, Santa Monica, California
Henry Chauncey, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey
Ralph McC. Chinn, University of Mississippi Medical Center, Dept. of Neurosurgery, Jackson, Mississippi

Paul Raymond Christensen, 8327 Georgetown Avenue, Los Angeles 45, California
Stephen C. Clark, 1759 La Paz Road, Altadena, California
Herbert H. Coburn, 71 Beacon Street, Boston 8, Massachusetts
Bernard P. Cohen, 312 Emerson Hall, Harvard University, Cambridge, Massachusetts
Jacob Cohen, 209 Concord Road, Yonkers, New York
Andrew Laurence Comrey, Jr., Dept. of Psychology, University of California, Los Angeles 24, California
Herbert S. Conrad, Research and Statistical Service, U. S. Office of Education, Washington 25, D. C.
Ellsworth Barrett Cook, U. S. N., Pharmacology Dept. N. M. R. I., Naval Medical Center, Bethesda 14, Maryland
William E. Cook, Alton State Hospital, Alton, Illinois
Clyde H. Coombs, Dept. of Psychology, University of Michigan, Ann Arbor, Michigan
Herman A. Copeland, 909 Atwood Road, Chesney Downs, Philadelphia 18, Pennsylvania
Walter Coppock, 1811 Vine Street, Berkeley 3, California
Francis G. Cornell, Engelhardt, Engelhardt, & Leggett, 221 W. 57th Street, New York 19, New York
Sarah Counts, 5143 S. Kenwood, Chicago 15, Illinois
John T. Cowles, Falk Clinic, University of Pittsburgh, Pittsburgh, Pennsylvania
Gertrude M. Cox, North Carolina State College of Agriculture and Engineering, Raleigh, North Carolina
H. M. Cox, University of Nebraska, Lincoln 8, Nebraska
Kenneth J. Cox, 35 W. Park Avenue, c/o West Hamilton P. O., Hamilton, Ontario, Canada
Clarke W. Crannell, 120 S. Main Street, Oxford, Ohio
William J. E. Crissy, 51-01 39th Avenue, Sunnyside, Long Island City, New York
Lee J. Cronbach, O. N. R. London, Box 39, Navy 100, F. P. O., New York
Dorothea J. Crook, Psychology Dept., Tufts College, Medford 55, Massachusetts
Elmer K. Culler, Laboratory of Psychology, River Campus, University of Rochester, Rochester 3, New York
E. E. Cureton, 1846 Prospect Place, S. E., Knoxville, Tennessee
Louise W. Cureton, 1846 Prospect Place S. E., Knoxville, Tennessee
Manuel Cynamon, Office of Research and Testing, Brooklyn College, Brooklyn 10, New York
W. Grant Dahlstrom, Dept. of Psychology, University of North Carolina, Chapel Hill, North Carolina
Karl M. Dallenbach, 2106 Meadowbrook Drive, Austin, Texas
M. B. Danford, Dept. of Biometrics, U. S. A. F., School of Aviation Medicine, Randolph Field, Texas
Melvin D. Davidoff, 10615 Dunkirk Drive, Silver Spring, Maryland
Frederick B. Davis, Educational Clinic, Hunter College, 695 Park Avenue, New York 21, New York
Paul C. Davis, c/o Roberts Wesleyan College, North Chili, New York
Hugh M. Davison, 419 Martin Terrace, State College, Pennsylvania
Walter L. Deemer, Jr., 3624 Appleton Street, N. W., Washington 8, D. C.
James W. Degan, Lincoln Laboratory, M. I. T., Lexington 73, Massachusetts
Robert G. Demaree, Armament Systems Personnel Research Laboratory, Lowry A. F. B., Denver, Colorado
Harvey F. Dingman, 2055 38th Street, S. E., Apt. 302, Washington 25, D. C.
Wilfrid J. Dixon, P. O. Box 5175, Eugene, Oregon
Jerome E. Doppelt, 18 Mansfield Road, White Plains, New York
Paul L. Dressel, Board of Examiners, Michigan State College, East Lansing, Michigan
Raymond J. Dry, Life Insurance Agency Management Association, 855 Asylum Avenue, Hartford 5, Connecticut

Philip H. DuBois, Dept. of Psychology, Washington University, St. Louis 5, Missouri
Robert D. Dugan, 3200 Allison Way, Louisville, Kentucky
Frank DuMas, Dept. of Psychology, Montana State University, Missoula, Montana
Jack W. Dunlap, Lockwood Road, Riverside, Connecticut
Marvin D. Dunnette, 209 S. E. Bedford, Minneapolis, Minnesota
Paul S. Dwyer, 403 Lenawee Drive, Ann Arbor, Michigan
Henry Dyer, 153 Jefferson Road, Princeton, New Jersey
Robert L. Ebel, Room 114 University Hall, Iowa City, Iowa
Harold Edgerton, 2 Nursery Road, New Canaan, Connecticut
Robert Edgren, 3410 Margarita Avenue, Oakland 5, California
Allen L. Edwards, Dept. of Psychology, University of Washington, Seattle 5, Washington
Daisy M. Edwards, 1 Oakfield Court, Queens Road, Weybridge, Surrey, England
Donald Ehrman, 158 Glenwood Avenue, Atherton, California
Daniel R. Embody, 24 Farmingdale Road, Fairview Knolls, Milltown, New Jersey
Max D. Engelhart, 5427 Greenwood Avenue, Chicago 15, Illinois
Donald Estavan, 223-C Halsey, Princeton, New Jersey
Katherine W. Estes, Dept. of Psychology, Indiana University, Bloomington, Indiana
Chester E. Evans, Employee Research Section, General Motors Corporation, Detroit 2, Michigan
Harold T. Fagin, 136 Jay Street, Johnson City, New York
Edmund W. Faison, Needham, Louis & Brorby, Inc., 135 La Salle Street, Chicago 3, Illinois
Chung-Teh Fan, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey
Mary E. Farabaugh, 4201 Massachusetts Avenue, N. W., Washington 16, D. C.
Leonard S. Feldt, Iowa Testing Programs, East Hall, Iowa City, Iowa
George A. Ferguson, Dept. of Psychology, Peterson Hall, McGill University, Montreal, Quebec, Canada
Samuel W. Fernberger, College Hall, University of Pennsylvania, Philadelphia 4, Pennsylvania
Leon Festinger, Dept. of Psychology, University of Minnesota, Minneapolis 14, Minnesota
Fred E. Fiedler, 240 Court A—P. G. U., Champaign, Illinois
Victor Fields, 7302 15th Avenue, Takoma Park, Maryland
Gordon Fifer, Apt. 61—610 W. 116th Street, New York 27, New York
Warren G. Findley, Educational Testing Service, Box 592, Princeton, New Jersey
John C. Flanagan, Dept. of Psychology, University of Pittsburgh, Pittsburgh 13, Pennsylvania
William J. Flynn, Jr., 1104 Milbourne Avenue, Flint 4, Michigan
Ronald C. Force, 4310 S. Lipscomb, Amarillo, Texas
John W. French, Educational Testing Service, Box 592, Princeton, New Jersey
Benjamin Fruchter, 1216 Norwood Road, Austin, Texas
Edmund F. Fuchs, 1806 Florin Street, Silver Spring, Maryland
Marguerite S. Gadel, Personnel Research Division, Prudential Insurance Co., Newark 1, New Jersey
N. L. Gage, Bureau of Educational Research, College of Education, University of Illinois, Urbana, Illinois
Henrietta L. Gallagher, 43 Wiggins Street, Princeton, New Jersey
Henry E. Garrett, Dept. of Psychology, Columbia University, New York 27, New York
Roman Gawkoski, 55 St. Andrews Place, Yonkers, New York
Richard H. Gaylord, 2550 Kemper Road, Cleveland 20, Ohio
Helen Hofer Gee, Dept. of Psychology, University of Minnesota, Minneapolis 14, Minnesota
Frederick Gehlmann, 1036 Shady Lane, Flowerfield, Lombard, Illinois
Charles P. Gershenson, 255 Fir Street, Park Forest, Illinois
Edwin E. Ghiselli, Dept. of Psychology, University of California, Berkeley, California
Jerome H. Gibson, 2957 Santa Clara Drive, Albuquerque, New Mexico

Wilfred A. Gibson, Personnel Research Branch, A.G.O., Department of the Army, Washington, D. C.
Robert Glaser, American Institute for Research, 410 Amberson Avenue, Pittsburgh 32, Pennsylvania
Goldine C. Gleser, 3604 Lansdowne Avenue, Cincinnati 36, Ohio
Alfred F. Glixman, Dept. of Psychology, University of Oklahoma, Norman, Oklahoma
Leon G. Goldstein, 7015 Freeport Street, Woodlawn, Hyattsville, Maryland
Maxim Goode, 6658 Whitley Terrace, Los Angeles 28, California
Leonard V. Gordon, U. S. Naval Personnel Research Unit, Bldg. 238 U. S. Navy Electronic Laboratory, San Diego 52, California
Mary Agnes Gordon, 545 Club Drive, San Antonio, Texas
Mordecai H. Gordon, Veterans Administration Hospital, Knoxville, Iowa
David A. Grant, Dept. of Psychology, University of Wisconsin, Madison 6, Wisconsin
Donald L. Grant, 17325 Fernway Road, Cleveland 20, Ohio
J. Douglas Grant, P. O. Box 365, Del Mar, California
Clifton W. Gray, Psychology Dept., University of Minnesota, Minneapolis 14, Minnesota
Bert F. Green, Jr., C-345, Box 73, Lexington 73, Massachusetts
Dorothy Green, U. S. Civil Service Commission, Rm. 133 Main Bldg., Washington 25, D. C.
Russel F. Green, 119 University Park, Rochester 20, New York
Richard B. Grekila, Box 143, Crenshaw, Pennsylvania
William W. Grings, Dept. of Psychology, University of Southern California, Los Angeles 7, California
J. P. Guilford, P. O. Box 1288, Beverly Hills, California
Harold O. Gulliksen, Box 592, Princeton, New Jersey
Hari C. Gupta, 5835 S. Kimbark Avenue, Chicago 37, Illinois
John W. Gustad, University Counseling Service, University of Maryland, College Park, Maryland
Isaiah Guttman, Dept. of Psychology, University of Minnesota, Minneapolis 14, Minnesota
Louis Guttman, The Israel Institute of Applied Social Research, Shell Building, Jerusalem, Israel
Raef K. Haddad, 1052 Brown, New York University, Washington Square, New York 3, New York
Elizabeth Hagen, 163 W. 17th Street, New York 11, New York
Harold Wesley Hake, Psychology Dept., Johns Hopkins University, Baltimore 18, Maryland
H. H. Harman, Rand Corporation, 1700 Main Street, Santa Monica, California
A. Edwin Harper, Jr., c/o Dr. P. C. Mahalanobis, Indian Statistical Institute, 204 Barrackpore Trunk Road, Calcutta 35, India
Mrs. Bertha P. Harper, Personnel Research Branch, T. A. G. O., Army, Washington 25, D. C. Attn: AGTL-T
Chester W. Harris, School of Education, University of Wisconsin, Madison, Wisconsin
Stephen Harrison, Research Laboratories, Kraft Foods Co., Glenview, Illinois
Charles M. Harsh, Human Factors Division, U. S. Navy Electronics Laboratory, San Diego 52, California
Edgar M. Haverland, 1325 W. Jackson Street, Kokomo, Indiana
Edward N. Hay, 245 Ogden Avenue, Swarthmore, Pennsylvania
Helen Heath, 7202 Bennett Avenue, Chicago 49, Illinois
Robert Arthur Hefner, Psychology Dept., University of Michigan, Ann Arbor, Michigan
Howard J. Heglin, 200 U. Flavet 3, Gainesville, Florida
Ruth Bishop Heiser, P. O. Box 47, Goshen, Kentucky
Carl E. Helm, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey
Gerald C. Helmstadter, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey
E. R. Henry, 428 Demarest Avenue, Oradell, New Jersey

Alfred F. Hertzka, 4719 Pescadero Avenue, San Diego 7, California
J. H. Hill, Psychology Dept., Washington University, St. Louis 5, Missouri
John R. Hills, Research Dept., Educational Testing Service, 20 Nassau Street, Princeton, New Jersey
Elmer D. Hinckley, Box 2007, University Station, Gainesville, Florida
Jerome E. Hirsch, 1127B Francisco Street, Berkeley 2, California
Lorain O. Hite, 3747 W. 59th Street, Apt. 4, Chicago 29, Illinois
E. Lee Hoffman, Urban Life Research Institute, Tulane University, New Orleans, Louisiana
Wayne H. Holtzman, Dept. of Psychology, University of Texas, Austin 12, Texas
David L. Holzman, 570 Lexington Avenue, New York 22, New York
Paul A. Horst, Dept. of Psychology, University of Washington, Seattle, Washington
Alston S. Householder, Oak Ridge National Laboratory, Oak Ridge, Tennessee
Cyril J. Hoyt, Bureau of Educational Research, University of Minnesota, Minneapolis, Minnesota
Edith M. Huddleston, Educational Testing Service, Princeton, New Jersey
Doncaster G. Humm, Box 1433 Del Valle Station, Los Angeles 15, California
Lloyd C. Humphreys, Personnel Research Laboratory, A.F.P.T.R.C., P. O. Box 1557, San Antonio, Texas
Cuthbert C. Hurd, International Business Machines Corporation, 590 Madison Avenue, New York 22, New York
Tarow Indow, Dept. of Psychology, Keio University; Mita, Shiba Minatuku; Tokyo, Japan
Walter Isaac, Dept. of Physiology & Biophysics, University of Washington, Seattle 5, Washington
Shinkuro Iwahara, Dept. of Psychology, Nara Women's University, Nara, Japan
Patricia Lee Jackson, 127 E. 35th Street, New York 16, New York
Rheem F. Jarrett, 1328 Josephine Street, Berkeley 3, California
Nathan Jaspen, National League of Nursing Education, 2 Park Avenue, New York 16, New York
T. N. Jenkins, 105-21 66th Avenue, Forest Hills, New York
William Leroy Jenkins, Dept. of Psychology, Lehigh University, Bethlehem, Pennsylvania
William A. Jenner, 7908 Kipling Parkway, S. E., Washington 28, D. C.
Cecil D. Johnson, 406 Fenwick Drive, Falls Church, Virginia
Duane M. Johnson, Cary Hall West, West Lafayette, Indiana
Palmer O. Johnson, Dept. of Education, University of Minnesota, Minneapolis 14, Minnesota
Woodbury Johnson, 8201 Tilbury Street, Bethesda, Maryland
Lyle V. Jones, Dept. of Psychology, University of Chicago, Chicago 37, Illinois
Clifford E. Jurgensen, Minneapolis Gas Co., Minneapolis, Minnesota
Lessing A. Kahn, 12021 Milton Street, Wheaton, Maryland
Henry F. Kaiser, 1818½ Tenth Street, Berkeley 2, California
Harry Kaplan, 3705 S Street S. E., Washington 20, D. C.
John E. Karlin, Bell Telephone Laboratories, Murray Hill, New Jersey
Bertram P. Karon, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey
Leo Katz, Dept. of Mathematics, Michigan State College, East Lansing, Michigan
Robert J. Keller, 222 Burton Hall, University of Minnesota, Minneapolis 14, Minnesota
H. Paul Kelley, Psychology Laboratory, School of Aviation Medicine, N. A. S., Pensacola, Florida
T. L. Kelley, 1611 Paterna Road, Santa Barbara, California
E. Lowell Kelly, 1115 Woodlawn Street, Ann Arbor, Michigan
Harold E. Kerber, P. O. Box 212, Chagrin Falls, Ohio
Norman Kettner, 1115 W. 31st Street, Los Angeles 7, California
Mary J. Kientzle, Dept. of Psychology, Washington State College, Pullman, Washington

Frank Killian, Jr., 3650 USAF Hospital, Box 630, Sampson AFB, New York
William E. Kline, The Choate School, Wallingford, Connecticut
Henry Klugh, 330 Melwood Avenue, Pittsburgh 13, Pennsylvania
Helen L. Koch, 1374 E. 57th Street, Chicago 37, Illinois
Leonard S. Kogan, Community Service Society of New York, 105 E. 22nd Street, New York 10, New York
Joseph F. Kubis, Fordham University, Graduate School, New York 58, New York
G. Frederic Kuder, College Station, Duke University, Durham, North Carolina
Albert K. Kurtz, Dept. of Psychology, University of Florida, Gainesville, Florida
Oliver L. Lacey, University of Alabama, University, Alabama
Thomas A. Lamke, Bureau of Research, Iowa State Teachers College, Cedar Falls, Iowa
Charles R. Langmuir, Psychological Corporation, 522 Fifth Avenue, New York 36, New York
Wilbur L. Layton, Student Counseling Bureau, 101 Eddy Hall, University of Minnesota, Minneapolis, Minnesota
Paul F. Lazarsfeld, Center for Advanced Study in the Behavioral Sciences, Box 97, Menlo Park, California
George S. Leavitt, Psychology Dept., Fresno State College, Fresno, California
Herbert B. Leedy, 536 Argyle Drive, Falls Church, Virginia
D. Welty Lefever, 4251 Don Felipe Drive, Los Angeles 8, California
John M. Leiman, Personnel Research Laboratory, Human Resources Research Center, Lackland AFB, San Antonio, Texas
Bernard Levenson, 535 W. 111th Street, New York 25, New York
Hollis M. Leverett, 21 Hampden Street, Wellesley, Massachusetts
Bernard Martin Levin, Dept. of Psychology, University of North Carolina, Chapel Hill, North Carolina
Richard S. Levine, Educational Testing Service, Princeton, New Jersey
Malcolm Lichtenstein, Code 2132 Navy Electronics Laboratory, San Diego 52, California
Fil dr. Rickard Lindahl, Övra Husargabar, Göteborg, Sweden
E. F. Lindquist, College of Education, State University of Iowa, Iowa City, Iowa
Kenneth B. Little, Psychology Dept., Stanford University, Stanford, California
Jane Loevinger, 7209 Pershing Avenue, St. Louis 5, Missouri
Eugene Robert Long, Jr., Psychology Dept., New West Bldg., University of North Carolina, Chapel Hill, North Carolina
Frederic Lord, 79 Dempsey Avenue, Princeton, New Jersey
Irving D. Lorge, 525 W. 120th Street, New York 27, New York
Maurice Lorr, 1521 Erskine Street, Washington 12, D. C.
Constance Lovell, Dept. of Psychology, University of Southern California, 3518 University Avenue, Los Angeles 7, California
Ardie Lubin, Psychology Dept., A.M.S.G.S., Walter Reed Army Medical Center, Washington 12, D. C.
Charles M. Lucas, 2666 South Street, Allentown, Pennsylvania
Bertha M. Luckey, Psychology Clinic, 1380 E. 6th Street, Cleveland 14, Ohio
George A. Lundberg, Dept. of Sociology, University of Washington, Seattle, Washington
Walter A. Lurie, National Community Relations Advisory Council, 9 E. 38th Street, New York 16, New York
Samuel B. Lyerly, 2405 Runlaw Road, N. W., Washington 7, D. C.
Howard B. Lyman, Dept. of Psychology, University of Cincinnati, Cincinnati 21, Ohio
R. R. Mackie, 3004 S. Grand Avenue, Los Angeles 7, California
Angus G. MacLean, 1461 Alvira Street, Los Angeles 35, California
Ernest Madril, 220 Test Squadron, HQ Continental Air Command, Mitchell AFB, New York

Benjamin Malzberg, Dept. of Mental Hygiene, State Office Bldg., Albany 1, New York
Melvin Marks, AGO, Personnel Research Section, Dept. of Army, Washington 25, D. C.
Frank Jones Massey, Jr., Mathematics Dept., University of Oregon, Eugene, Oregon
William G. Matheny, 1903 Tyndall Drive, Panama City, Florida
Luzelle Denton Mays, 1534 16th Road N., Apt. 12, Arlington, Virginia
Harold L. McAdoo, 914 Barrett Road, Falls Church, Virginia
Robert L. McCornack, Dept. of Psychology, Wayne University, Detroit 1, Michigan
William J. McGill, Rm. 22A-118A, Massachusetts Institute of Technology, Cambridge 39, Massachusetts
Richard McHugh, Dept. of Psychology, Iowa State College, Ames, Iowa
Kenneth Floyd McLaughlin, Test Service Bureau, Florida State University, Tallahassee, Florida
Walter J. McNamara, IBM Corporation, 590 Madison Avenue, New York, New York
Quinn McNemar, Dept. of Psychology, Stanford University, Stanford, California
Louis L. McQuitty, Dept. of Psychology, University of Illinois, Urbana, Illinois
Francis F. Medland, Marine Corps (Code DC), Rm. 4030 Arl. Annex, Washington 25, D. C.
Donald M. Medley, Div. of Teacher Education, College of the City of New York, 500 Park Avenue, New York 22, New York
Paul E. Meehl, 112 Psychology Bldg., University of Minnesota, Minneapolis 14, Minnesota
J. J. Mellinger, Psychometric Laboratory, University of North Carolina, Chapel Hill, North Carolina
Peter F. Merenda, 1701 70th Street, Kenosha, Wisconsin
Charlotte Levy Mernyk, 1602 E. 18th Street, Brooklyn 30, New York
Samuel J. Messick, The Menninger Foundation, Topeka, Kansas
Charles A. Metzner, Rm. 2539, School of Public Health, University of Michigan, Ann Arbor, Michigan
Milton Meux, 1011 W. Green Street, Urbana, Illinois
Lorenz A. Meyer, 1814 Kanawha Street, Adelphi, Maryland
William Burton Michael, Test Bureau, University of Southern California, University Park, Los Angeles 7, California
John E. Milholland, Psychology Dept., University of Michigan, Ann Arbor, Michigan
George Armitage Miller, 14 Barberry Road, Lexington 73, Massachusetts
John A. Modrick, Washington and Jefferson College, Washington, Pennsylvania
William G. Mollenkopf, 112 Cedar Lane, Princeton, New Jersey
William T. Moran, Young & Rubicam, 285 Madison Avenue, New York, New York
Robert E. Morin, Dept. of Psychology, University of Texas, Austin, Texas
Donald R. Morrison, 219 Cherry Street, Carlisle, Pennsylvania
Edward John Morrison, 1271 Xapary, Denver 8, Colorado
Frederick Mosteller, 324 Emerson Hall, Harvard University, Cambridge 38, Massachusetts
John F. Muldoon, 1113 Glen Eagle Road, Baltimore 12, Maryland
Carl Murchison, The Journal Press, 2 Commercial Street, Provincetown, Massachusetts
Paul M. Neurath, Queens College, Flushing, New York
Wilson L. Newman, 5539 S. Kenwood Avenue, Chicago 37, Illinois
William T. Nicholas, No. 7 Parkview Court, Pocatello, Idaho
Warren T. Norman, 1929 2nd Street, N. E., Minneapolis 18, Minnesota
Raymond C. Norris, Dept. of Psychology, George Peabody College for Teachers, Nashville 4, Tennessee
James A. Norton, Jr., Statistics Laboratory, Purdue University, Lafayette, Indiana
Ewald R. Nyquist, R. F. D. No. 1, Ravena, New York
Francisco del Olmo, Industrial Relations Dept., Creole Petroleum Corporation, Apartado 889, Caracas, Venezuela
Willard C. Olson, School of Education, University of Michigan, Ann Arbor, Michigan

George N. Ornstein, Det. No. 6 (Interceptor Pilot Research Laboratory), Tyndall AFB, Panama City, Florida
David B. Orr, 509 W. 121st Street, New York 27, New York
Bernice Orshansky, 1695 Grand Concourse, New York 53, New York
Hobart G. Osburn, 8905 Kimes, Silver Spring, Maryland
Herbert Ozur, 7315 17th Avenue, West Hyattsville, Maryland
James W. Parker, Jr., Personnel Research Institute, Western Reserve University, Cleveland 6, Ohio
Siroon Pashalian, 65 W. 181st Street, New York 53, New York
Donald C. Pelz, 909 E. Huron Street, Ann Arbor, Michigan
Lee Edward Paul, Dept. of Psychology, University of North Carolina, Chapel Hill, North Carolina
Robert Perloff, Science Research Associates, 57 W. Grand Avenue, Chicago 10, Illinois
Donald A. Peterson, 855 Asylum Avenue, Hartford 5, Connecticut
Robert M. Peterson, Brigham Young University, Provo, Utah
Shailer A. Peterson, American Dental Association, 222 E. Superior Avenue, Chicago 11, Illinois
B. Roger Philip, 323 Annette Street, Toronto, Canada
H. T. H. Piaggio, Dept. of Mathematics, University College, University Park, Nottingham, England
Henri Pieron, Institut de Psychologie à la Sorbonne, Rue des Ecoles, Paris 5, France
A. E. G. Pilliner, Moray House, Edinburgh 8, Great Britain
Joseph A. Pittman, North Carolina College at Durham, Durham, North Carolina
Lynnette Birdsall Plumlee, Box 592, Princeton, New Jersey
A. Terrence Polin, 2877 Grand Concourse, Bronx 68, New York
Robert Radlow, 420 S. Burrowes Street, State College, Pennsylvania
James A. Rafferty, 1425 Valleycrest Boulevard, Falls Church, Virginia
John H. Rapparlie, R. F. D. No. 1, Oak Harbor, Ohio
Philburn Ratoosh, Dept. of Psychology, 404 University Hall, Ohio State University, Columbus 10, Ohio
James J. Reilly, 160 Brook Street, Manchester, New Hampshire
William Reitz, Educational Evaluation, Statistics & Research, Wayne University Student Center, Detroit 1, Michigan
Mary E. Reuder, Apt. 3V, 82-46 135th Street, Kew Gardens 35, New York
Yvonne A. Roach, 25-41 97th Street, East Elmhurst 69, New York
Marguerite E. Robertson, 221 Adams Street, Lakewood 5, Rhode Island
William L. Robinson, Jr., 1542-A Brackenridge Apts., Austin, Texas
Milton L. Rock, Edward N. Hay & Associates, Inc., 1523 Walnut Street, Philadelphia 2, Pennsylvania
Merrill Roff, University of Minnesota, Institute of Child Welfare, Minneapolis 14, Minnesota
Robert C. Rogers, R. F. D. No. 1, Princeton, New Jersey
Benjamin Rosner, 316 Bay 20th Street, Brooklyn 14, New York
Paul F. Ross, 3167 N. High Street, Columbus, Ohio
Thomas C. Rowan, Rand Corporation, 1905 Armacost, Los Angeles 25, California
Joseph R. Royce, Dept. of Psychology, University of Redlands, Redlands, California
Philip J. Rulon, Peabody House, 13 Kirkland Street, Cambridge 38, Massachusetts
J. Francis Rummel, U.S.O.M./Egypt, c/o American Embassy, Cairo, Egypt
Edward A. Rundquist, 1405 Summit Avenue, Alexandria, Virginia
Jean M. Rutherford, Box 24, Arizona State College, Tempe, Arizona
T. A. Ryan, Dept. of Psychology, Cornell University, Ithaca, New York
David G. Ryans, Dept. of Education, University of California, Los Angeles 24, California

Robert Sadacca, 107 E. 88th Street, New York 28, New York
Edward Sait, Rensselaer Polytechnic Institute, Personnel Counseling & Testing Laboratory, 8th Street & Sage Avenue, Troy, New York
James M. Sakoda, Dept. of Psychology, University of Connecticut, Storrs, Connecticut
Richard T. Sampson, Personnel Staff, Statistical Service, Farm Bureau Insurance Co., 246 N. High Street, Columbus, Ohio
Harry Saslow, Dept. of Psychology, University of Pittsburgh, Pittsburgh 13, Pennsylvania
David R. Saunders, Educational Testing Service, Princeton, New Jersey
Douglas E. Scates, American Social Hygiene Association, 1790 Broadway, New York 19, New York
Willis C. Schaefer, The Rand Corporation, 1700 Main Street, Santa Monica, California
W. Warner Schaie, Dept. of Psychology, University of Washington, Seattle 5, Washington
Kenneth F. Schenkel, 5916-22nd Place S. E., Washington 21, D. C.
Lowell Schipper, 259 Natatorium, Laboratory of Aviation Psychology, Ohio State University, Columbus 10, Ohio
John Schmid, Personnel Research Laboratory, Human Resources Research Center, Lackland AFB, San Antonio, Texas
William B. Schrader, Educational Testing Service, Princeton, New Jersey
Douglas Schultz, Jericho Manor, Apt. 130, Jenkintown, Pennsylvania
Robert W. Scollay, 19 S. Franklin Street, Madison 3, Wisconsin
Saul B. Sells, 306 Abiso Avenue, San Antonio 9, Texas
Marion F. Shaycoft, 406 N. Neville Street, Pittsburgh 13, Pennsylvania
Joseph Russell Sherlock, 2685 Briggs Avenue, New York 58, New York
R. H. Shevenell, Institute of Psychology, University of Ottawa, Ottawa, Canada
Nathan Shock, Baltimore City Hospitals, Unit on Gerontology, Baltimore 24, Maryland
Meyer Shultz, 537 Peabody Street N. W., Washington, D. C.
Herbert S. Sichel, Operational Research Bureau, P. O. Box 10238, Johannesburg, South Africa
Paul S. Siegel, Dept. of Psychology, University of Alabama, University, Alabama
Sidney Siegel, Dept. of Psychology, Pennsylvania State University, University Park, Pennsylvania
George B. Simon, 808 S. Grove Street, Urbana, Illinois
Roy G. Simpson, 1859 Page Avenue, East Cleveland, Ohio
J. E. Keith Smith, 112 Longwood Avenue, Brookline 46, Massachusetts
Max Smith, 453 E. 96th Street, Brooklyn 12, New York
Herbert Solomon, Teachers College, Columbia University, New York 27, New York
Joan L. Sonnenschein, 74 Walnut Street, Somerville, Massachusetts
Thomas B. Sprecher, VF B, Norwich Road, College Park, Maryland
Richard E. Stafford, 2012 Delancey Street, Philadelphia, Pennsylvania
John M. Stalnaker, 1075 Elm Street, Winnetka, Illinois
Julian C. Stanley, Jr., Eckhart Hall, University of Chicago, Chicago 37, Illinois
Artie John Stauffer, Box 6034, San Diego 6, California
John E. Stecklein, 211 Burton Hall, University of Minnesota, Minneapolis, Minnesota
Frederick F. Stephan, 105 FitzRandolph Road, Princeton, New Jersey
Theodor D. Sterling, Dept. of Mathematics, University of Alabama, P. O. Box 3187, Tuscaloosa, Alabama
Jack J. Sternberg, 8702 Gilbert Place, Takoma Park, Washington 12, D. C.
Naomi Stewart, 13 Doering Way, Cranford, New Jersey
Glen F. Stice, Educational Testing Service, Princeton, New Jersey
Edward K. Strong, Jr., 672 Mirada Avenue, Stanford, California
Frances Swineford, Educational Testing Service, Princeton, New Jersey
Nancy A. Symons, 401 Washington Avenue, Elyria, Ohio

Arthur T. Tait, 5523 Rumsey Drive, Riverside, California
Alfonso Tallet-Grenier, Universidad de Oriente, Santiago de Cuba, Cuba
Maurice M. Tatsuoka, 64 Wendell Street, Cambridge, Massachusetts
Calvin W. Taylor, Dept. of Psychology, University of Utah, Salt Lake City, Utah
Erwin K. Taylor, 1215 Terminal Tower, Cleveland 13, Ohio
Harold C. Taylor, W. E. Upjohn Institute for Community Research, 709 S. Westnedge Avenue, Kalamazoo 4, Michigan
Newell E. Taylor, 600 Pelham Road, New Rochelle, New York
Lewis M. Terman, 761 Dolores Street, Stanford University, Stanford, California
Kenneth F. Thomson, 3801 Hayfield Place, Alexandria, Virginia
Robert L. Thorndike, Teachers College, Columbia University, New York 27, New York
Robert M. Thrall, Dept. of Mathematics, University of Michigan, Ann Arbor, Michigan
Thelma G. Thurstone, Psychometric Laboratory, University of North Carolina, Chapel Hill, North Carolina
David V. Tiedeman, Peabody House, 13 Kirkland Street, Cambridge 38, Massachusetts
Kalman Tillem, 2968 Calvert, Detroit, Michigan
Helen Tomlinson, 354 Broadview, San Antonio, Texas
Herbert A. Toops, 1430 Cambridge Boulevard, Columbus 8, Ohio
Warren S. Torgerson, 36 Genetti Circle, Bedford, Massachusetts
Arthur E. Traxler, 21 Audubon Avenue, New York 32, New York
David K. Trites, 725 Patterson Avenue, San Antonio, Texas
William F. Tronca, 2506 30th Street, Seattle 44, Washington
William Trousdale, 1206 Fourth St. S. E., Minneapolis, Minnesota
James B. Trump, 2814 74th Avenue, Apt. 202, Hyattsville, Maryland
Anthony C. Tucker, 427 Boynton Avenue, Berkeley 7, California
Ledyard R Tucker, 102 Cedar Lane, Princeton, New Jersey
Dik W. Twedt, Needham, Louis & Brorby, Inc., 135 S. La Salle Street, Chicago 3, Illinois
Fred T. Tyler, Haviland Hall, University of California, Berkeley, California
Ralph W. Tyler, Center for Advanced Study in the Behavioral Sciences, 202 Junipero Serra, Stanford, California
E. J. Uhlaner, 5430 N. 20th Street, Arlington 5, Virginia
C. C. Upshall, 21 Eastview Terrace, Pittsford, New York
Harry S. Upshaw, Box 1552, University of Alabama, Tuskaloosa, Alabama
Steven G. Vandenberg, Institute of Human Biology, 1135 E. Catherine, Ann Arbor, Michigan
Neil J. Van Steenberg, 5556 N. 16th Street, Arlington, Virginia
Marvin J. Van Wagenen, University of Minnesota, Minneapolis 14, Minnesota
Manmohan Varma, University Training College, Nagpur (MP), India
C. L. Vaughn, Psychological Corporation, 522 Fifth Avenue, New York 36, New York
Mary Vermillion, 5140 Kenwood Avenue, Chicago 15, Illinois
David F. Votaw, Box 587, San Marcos, Texas
William B. Walker, General Electric Co., Public Relations Research Service, 570 Lexington Avenue, New York, New York
M. J. Wantman, Catherine Strong Hall, University of Rochester, Rochester 3, New York
Joe H. Ward, Jr., 315 Palm Drive, San Antonio, Texas
Neil D. Warren, 3827 Olympiad Drive, Los Angeles 43, California
Richard W. Watkins, E.T.S., 20 Nassau Street, Princeton, New Jersey
Herbert B. Weaver, Dept. of Psychology, University of Hawaii, Honolulu, Hawaii
Harold Webster, Vassar College, Poughkeepsie, New York
Arla Weinert, 905 Fairview Avenue, Apt. 3, Washington 12, D. C.
Ralph Weintraub, 466 87th Street, Brooklyn 9, New York
George Welsh, Psychology Dept., University of North Carolina, Chapel Hill, North Carolina

Alexander G. Wesman, 84 Ralph Avenue, White Plains, New York
Risdon John Westen, 1402 Briarcliff Drive, Urbana, Illinois
Benjamin W. White, Human Research Unit No. 2, P. O. Box 446, Fort Ord, California
Dean K. Whitla, Office of Tests, 11 University Hall, Cambridge 38, Massachusetts
Alfred G. Whitney, Life Insurance Agency Management Association, 855 Asylum Avenue, Hartford 5, Connecticut
S. S. Wilks, Dept. of Mathematics, Princeton University, Princeton, New Jersey
Richard A. Willard, 7-A William James Hall, Cambridge 38, Massachusetts
Louis P. Willemin, 2425 14th Street N. E., Washington 18, D. C.
Richard H. Willis, Group Behavior Laboratory, University of Wisconsin, Madison, Wisconsin
Clark L. Wilson, Jr., 3004 S. Grand Avenue, Los Angeles 7, California
Ben James Winer, Dept. of Psychology, Purdue University, Lafayette, Indiana
John F. Winne, Dept. of Psychology, Veterans Administration Hospital, Northampton, Massachusetts
Herman O. A. Wold, Institute of Statistics, Skolgatan 1, Uppsala, Sweden
Dael L. Wolfle, 2801 Rittenhouse Street N. W., Washington 15, D. C.
Ben D. Wood, Columbia University, New York 27, New York
Max A. Woodbury, 707 22nd Street, N. W., Washington 7, D. C.
Calvin Wright, 2918 Fuhrman Avenue, Seattle 2, Washington
J. Wayne Wrightstone, 21 Hickory Road, Summit, New Jersey
C. F. Wrigley, Research Associate, Dept. of Psychology, University of Illinois, Urbana, Illinois
James K. Yarnold, Apt. 103, 5482 S. Greenwood Avenue, Chicago, Illinois
Matsusaburo Yokoyama, Dept. of Psychology, Keio University, Mita, Tokyo, Japan
Irving Yudelowitz, 11 Midwood Street, Brooklyn 25, New York
Virginia Zachert, R. F. D. 2, Norman Park, Georgia
Robert C. Ziller, Crew Research Laboratory, A.F.P.&T.R.C., Randolph Field, Texas
Joseph Zubin, 722 W. 168th Street, New York, New York

PSYCHOMETRIKA—VOL. 21, NO. 2
JUNE, 1956

OPTIMAL TEST LENGTH FOR MAXIMUM ABSOLUTE PREDICTION*

PAUL HORST AND CHARLOTTE MACEWAN
UNIVERSITY OF WASHINGTON

The concepts of multiple differential prediction and multiple absolute prediction are developed in earlier papers (2, 3). The problem of determining the optimal distribution of testing time for multiple differential prediction has been previously considered (4). This paper develops an analogous procedure for multiple absolute prediction. A numerical example illustrating the procedure is presented. The mathematical rationale underlying the procedure is given.

I. *The Problem*

A technique was presented in (3) for selecting from a large number of predictor variables the subset of specified size which would have the highest absolute prediction efficiency for a given set of criterion variables. In (2), an analogous procedure was developed for selecting the subset which would most efficiently predict the multiple criteria differentially. In each of these cases, efficiency was defined in terms of the accuracy of prediction. In (2), differential prediction efficiency was defined in terms of the accuracy with which differences between all possible pairs of criterion measures could be predicted. An appropriate index of the prediction efficiency of a selected battery was shown to be the difference between the average variance of the predicted criteria and their average covariance. This index was designated by ϕ; the larger the value of ϕ, the greater the differential prediction efficiency of the battery.

In the case of multiple absolute prediction (3), efficiency was defined in terms of the accuracy with which all the criteria could be predicted, regardless of the extent to which the selected battery would differentiate among them. The index of absolute prediction efficiency of the selected battery was taken as the sum of the variances of the predicted criteria, regardless of their covariance; the larger this sum, designated by λ, the greater the absolute prediction efficiency of the battery.

*This research was carried out under Contract Nonr-477(08) between the University of Washington and the Office of Naval Research. The computations were carried out by Robert Dear and Donald Mills. Much credit is due the typist, Elizabeth Cross. Supervision of both computational and editorial activities was provided by William Clemans. To each of these able contributors we are deeply grateful.

For both types of multiple prediction it was assumed that intercorrelations between the potential predictors were available, as well as satisfactory estimates of the correlations between each criterion variable and the potential predictors. It was further assumed that predictors and criteria were in standard measures, and that the predicted criteria were the least square estimates. As was pointed out in (3), the essential difference between multiple absolute and multiple differential prediction is that the respective sets of predictors selected will differ. The two sets may show varying degrees of overlap; the extent of overlap depends upon the degree of correlation among the criterion variables and upon the original group of potential predictors from which the two subsets were selected.

Both methods referred to tacitly assume that all potential predictors take the same amount of administration time, so that all subsets of the same size would take equal administration time. This will not usually be the case. The problem may be approached in a more general way by starting with a given battery of predictor variables and determining how, for a specified over-all testing time, the individual test time-limits should be altered in order to maximize the index of absolute prediction efficiency.

For the case of a single criterion, a method is available (1) for determining optimal distribution of over-all administration time for a given battery of predictors. In (4) the method is modified and generalized for differential prediction involving multiple criteria. In this article the procedure developed in (1) is extended to the case of multiple absolute prediction.

As published, the method presented in (1) for solving for optimal test length assumes that the regression weights for the tests of optimal length are all positive. Otherwise the optimal solution could lead to the unacceptable solution where some of the optimal test lengths are negative. The extension of the technique in this article provides a computational solution which cannot yield negative test length. Since the method in (1) is a special case of the technique presented in the present article the more general method may also be used to avoid negative test lengths for the case of a single criterion. However, an iterative solution is required, whereas in the former case an exact solution was indicated.

In the present paper, as in the cases previously presented, it is assumed that intercorrelation, validity, and reliability data are available for predictors of arbitrary length. Testing time is taken to be the time actually allotted the examinee for taking the test. Any alteration in testing time implies a corresponding alteration in the number of items. Consequently the terms "testing time" and "test length" are used synonymously.

The method will first be described and illustrated by a numerical example. Following this, the mathematical rationale will be presented.

II. *Numerical Example*

The data used for this example are the same as those presented in (4). The predictor variables are:

(1) Guilford-Zimmerman Aptitude Survey Part I, Verbal Comprehension

(2) Guilford-Zimmerman Aptitude Survey Part III, Numerical Operations

(3) Guilford-Zimmerman Aptitude Survey Part VII, Mechanical Knowledge

(4) A. C. E. Psychological Examination, Quantitative Reasoning

(5) A. C. E. Psychological Examination, Linguistic Reasoning

(6) Cooperative English Test (Form OM), Usage

The matrix of test intercorrelations with reliabilities in the diagonal is given in Table 1. The criterion variables are grade point averages in each of ten college areas. The matrix of validity coefficients is given in Table 2.

It is evident that the correlations of variable 3 with the criterion are all small, four of them being negative. In general the chief justification for including such a variable would be that it might serve as a suppressor variable, i.e., a variable which suppresses invalid systematic variance in the predictor variables.

The original test lengths may be seen in row 1 of Table 3. The over-all testing time for the tests of arbitrary length is 142 minutes. We assume that this time is to be cut in half so that the over-all testing time is 71 minutes. The problem is to determine the time to be allotted to each test so as to maximize the index of absolute prediction efficiency.

The traditional assumptions are used here as in (1) with respect to the effect of test length on correlation, and will not be repeated. As in (4), the method for solving for the new test lengths involves a series of successive approximations. The computational procedure to be described consists of the same sequence of operations as those given in (4), the difference being that in the current procedure the matrix of validity coefficients is used, whereas in (4) the deviation form of these coefficients was required.

1. The first computational step is the calculation of the elements for a diagonal matrix Δ, seen in Table 3, row 3, labelled $1'\Delta$. Row 1 of this table gives the original lengths of tests; the elements in row 2 were obtained by subtracting the reliability of the indicated test from unity. The elements of Δ in row 3 are obtained by multiplying the element in row 1 by the corresponding element in row 2. Thus for the first element we have $\Delta_1 = 25(.080) = 2.000$.

2. A first approximation is now required for the altered test lengths. We assume the new test lengths to be proportional to the original test lengths.

TABLE 1

R Matrix of Predictor Intercorrelations with Reliabilities Substituted for Unities in the Diagonal:

$$R = r - D_u$$

		1	2	3	4	5	6	Σ
1	G-Z I	.920	.159	.152	.281	.763	.515	2.790
2	G-Z III	.159	.920	.003	.369	.292	.243	1.986
3	G-Z VII	.152	.003	.920	.200	.142	-.150	1.267
4	A.C.E.-Q	.281	.369	.200	.820	.549	.426	2.645
5	A.C.E.-L	.763	.292	.142	.549	.830	.628	3.204
6	English	.515	.243	-.150	.426	.628	.860	2.522
Σ		2.790	1.986	1.267	2.645	3.204	2.522	14.414

TABLE 2

The r_c' Matrix of Validity Coefficients

		1 G-Z I	2 G-Z III	3 G-ZVII	4 ACE-Q	5 ACE-L	6 English	Σ
1	Anthropology	.370	.177	.091	.294	.341	.357	1.630
2	Chemistry	.317	.274	.016	.309	.364	.399	1.679
3	Economics	.339	.211	.008	.241	.334	.323	1.456
4	English	.526	.247	-.075	.262	.488	.524	1.972
5	Foreign Lang.	.295	.287	-.156	.200	.232	.426	1.284
6	Geology	.184	.140	.094	.170	.229	.214	1.031
7	History	.379	.169	-.001	.182	.373	.336	1.438
8	Mathematics	.287	.348	-.088	.350	.336	.401	1.634
9	Psychology	.440	.170	.096	.285	.409	.403	1.803
10	Zoology	.336	.216	.031	.318	.345	.351	1.597
Σ		3.473	2.239	.016	2.611	3.451	3.734	15.524

TABLE 3

Computation of $1'\Delta$, $1'D_{b_1}^{-1}$ and $1'\Delta D_{b_1}^{-1}$

First approximation: $1'D_{b_1} = \frac{1}{2}1'D_a$

		1	2	3	4	5	6	Ck	Σ
1	$1'D_a$	25.0	9.0	30.0	23.0	15.0	40.0		142.0
2	$1'D_u$	.080	.080	.080	.180	.170	.140		
3	$1'\Delta$	2.000	.720	2.400	4.140	2.550	5.600		17.410
4	$1'D_{b_1}=.5(1'D_a)$	12.5	4.5	15.0	11.5	7.5	20.0	71.0	71.0
5	$1'D_{b_1}^{-1}$	.080	.222	.067	.087	.133	.050		
6	$1'\Delta D_{b_1}^{-1}$	.160	.160	.161	.360	.339	.280		1.460

Therefore, as a first approximation to the new test lengths we take one half the original test lengths. Row 4 in Table 3 is obtained as one half of row 1.

3. Calculate the reciprocal for each of the D_b elements. These are shown in row 5 of Table 3.

4. Multiply each Δ value in row 3 of Table 3 by the corresponding value in row 5. The products are entered in row 6 of Table 3. For example, the first element is .160 = 2.000(.080).

5. Next, the elements calculated in step 4 are added to the corresponding diagonal elements of Table 1, and the table is copied into the upper left quadrant of Table 4. The first diagonal element is 1.080 = .160 + .920. Note that the elements below the diagonal are not copied in. The upper right section of Table 4 is the matrix r_e, the transpose of Table 2.

6. We next calculate a matrix L_1 by premultiplying the matrix r_e by the inverse of the matrix in the upper left quadrant of Table 4. The computations of the forward solution are given in the two lower quadrants of Table 4 and in Table 5. The back solution is given in Table 6, in which the transpose of the matrix L_1 appears in the upper left corner. The procedure for multiplying a matrix by the inverse of a symmetric matrix is outlined in (5).

7. The second approximation to the new test lengths is computed in the lower section of Table 6 as follows:

Row a consists of the sums of squares of the column elements of the L_1' matrix. For example, the first element of row a, namely, .441, is the sum of squares of the first ten elements in column 1 of Table 6. Row b is copied from row 3 of Table 3. Row c consists of products of corresponding elements in rows a and b. For example, for the first element, .882 = .441(2.000).

Row d is obtained by taking the square root of the corresponding element in row c. For example, .939 = $\sqrt{.882}$. Row e, a check upon the computation of row d, consists of the squares of corresponding elements in row d. Row e, then, should be the same as row c. The value computed to the right of row e, and labelled s_1, is obtained by dividing the new over-all testing time, 71 minutes, by 3.930, the sum of elements in row d. Thus the value of s_1 is 18.066.

Row f is obtained by multiplying each element of row d, including the summation element, by s_1. For example, the first element obtained is 16.964 = .939(18.066). This row gives the second approximation to the new test lengths and its sum should equal the new over-all testing time.

Row g is obtained by dividing each element in row b by the corresponding element in row f. For example, the first element is .118 = 2.000/16.964.

Row h is obtained by adding the elements in row g to the corresponding reliabilities as given in the diagonal of Table 1. For example, the first element in row h is .118 + .920 = 1.038.

8. A new L_2 matrix is computed by repeating steps 5 and 6, and using the elements of row h of Table 6 in the diagonal positions of Table 1. The L_2 matrix may be seen in transposed form in Table 7, rows 1 through 10.

TABLE 4

Computation of $L_1 = (R+\Delta D_{b_1}^{-1})^{-1} r_c$ Forward Solution

Recip	1	1A	2A	3A	4A	5A	6A	1B	2B	3B	4B	5B	6B	7B	8B	9B	10B	C	S
	1A	1.080	.159	.152	.281	.763	.515	.370	.317	.339	.526	.295	.184	.379	.287	.440	.336	6.423	
	2A		1.080	.003	.369	.292	.243	.177	.274	.211	.247	.287	.140	.169	.348	.170	.216	4.385	
	3A			1.081	.200	.142	-.150	.091	.016	.008	-.075	-.156	.094	-.001	-.088	.096	.031	1.444	
	4A				1.180	.549	.426	.294	.309	.241	.262	.200	.170	.182	.350	.285	.318	5.616	
	5A					1.169	.628	.341	.364	.334	.488	.232	.229	.373	.336	.409	.345	6.994	
	6A						1.140	.357	.399	.323	.524	.426	.214	.336	.401	.403	.351	6.536	
	C	2.950	2.146	1.428	3.005	3.543	2.802	1.630	1.679	1.456	1.972	1.284	1.031	1.438	1.634	1.803	1.597	31.398	31.398
.926	1B	1.080	.159	.152	.281	.763	.515	.370	.317	.339	.526	.295	.184	.379	.287	.440	.336	6.423	6.423
.946	2B		1.057	-.019	.328	.180	.167	.123	.227	.161	.170	.244	.113	.113	.306	.105	.167	3.441	3.442
.944	3B			1.059	.166	.038	-.220	.041	-.025	-.037	-.146	-.193	.070	-.052	-.123	.036	-.013	.600	.601
1.021	4B				.979	.289	.275	.153	.160	.109	.095	.078	.076	.057	.200	.132	.181	2.785	2.784
1.953	5B					.512	.162	.012	.055	.036	.064	-.034	.055	.071	.026	.040	.026	1.025	1.025
1.441	6B						.694	.123	.144	.086	.169	.195	.084	.088	.126	.134	.103	1.947	1.946

TABLE 5

Computation of $L_1 = (R+\Delta D_{b_1}^{-1})^{-1} r_c$ (continued)

1	1A	2A	3A	4A	5A	6A	1B	2B	3B	4B	5B	6B	7B	8B	9B	10B	C	S
1A	1.000																	
2A		1.000																
3A			1.000															
4A				1.000														
5A					1.000													
6A						1.000												
1B	-1.000	-.147	-.141	-.260	-.707	-.477	-.343	-.294	-.314	-.487	-.273	-.170	-.351	-.266	-.407	-.311	-5.948	-5.948
2B		-1.000	.018	-.310	-.170	-.158	-.116	-.215	-.152	-.161	-.231	-.107	-.107	-.289	-.099	-.158	-3.255	-3.255
3B			-1.000	-.157	-.036	.208	-.039	.024	.035	.138	.182	-.066	.049	.116	-.034	.012	- .566	- .568
4B				-1.000	-.295	-.281	-.156	-.163	-.111	-.097	-.080	-.078	-.058	-.204	-.135	-.185	-2.843	-2.843
5B					-1.000	-.316	-.023	-.107	-.070	-.125	.066	-.107	-.139	-.051	-.078	-.051	-2.002	-2.001
6B						-1.000	-.177	-.208	-.124	-.244	-.281	-.121	-.127	-.182	-.193	-.148	-2.806	-2.805

TABLE 6

Computation of $L_1 = (R+\Delta D_{b_1}^{-1})^{-1} r_c$ Back Solution

i		1A	2A	3A	4A	5A	6A	1B	2B	3B	4B	5B	6B	7B	8B	9B	10B	Check	
1		.235	.059	.059	.116	-.033	.177	-1										.002	0
2		.120	.147	.003	.092	.041	.208		-1									-.001	0
3		.203	.106	-.021	.067	.031	.124			-1								-.001	0
4		.330	.108	-.091	.014	.048	.244				-1							-.001	0
5		.225	.196	-.125	.047	-.155	.281					-1						.001	0
6		.035	.070	.085	.024	.069	.121						-1					-.001	0
7		.215	.072	-.025	-.007	.099	.127							-1				.001	0
8		.127	.212	-.102	.155	-.007	.182								-1			.001	0
9		.268	.043	.060	.076	.017	.193									-1		-.005	0
10		.188	.090	-.004	.142	.004	.148										-1	.001	0
a	$1'D_{L_1L_1'}$	.441	.151	.050	.079	.045	.352	Σ	Check										
b	$1'\Delta$	2.000	.720	2.400	4.140	2.550	5.600												
c	$1'D_{L_1L_1'}\Delta$	.882	.109	.120	.327	.115	1.971												
d	$1'(D_{L_1L_1'}\Delta)^{\frac{1}{2}}$	.939	.330	.346	.572	.339	1.404	3.930	3.930										
e	$1'(D_{L_1L_1'}\Delta)^{\frac{1}{2}}(D_{L_1L_1'}\Delta)^{\frac{1}{2}}$	.882	.109	.120	.327	.115	1.971												
f	$1'D_{b_2} = 1'(D_{L_1L_1'}\Delta)^{\frac{1}{2}}s_1$	16.964	5.962	6.251	10.334	6.124	25.365	71.000	71.000										
g	$1'\Delta D_{b_2}^{-1}$	.118	.121	.384	.400	.416	.221	1.660											
h	$1'D_R + 1'\Delta D_{b_2}^{-1}$	1.038	1.041	1.304	1.220	1.246	1.081	6.930	6.930										

$$s_1 = \frac{T}{1'(D_{L_1L_1'}\Delta)^{\frac{1}{2}}1} = \frac{71}{3.930} = 18.066$$

TABLE 7

Computation of $L_2 = (R + \Delta D_{b_2}^{-1})^{-1} r_c$ Back Solution

1	1A	2A	3A	4A	5A	6A	1B	2B	3B	4B	5B	6B	7B	8B	9B	10B	Check	
1	.247	.061	.049	.110	-.040	.186	-1										-.002	0
2	.126	.153	.008	.085	.026	.227		-1									-.001	0
3	.216	.111	-.015	.062	.016	.135			-1								-.002	0
4	.350	.113	-.071	.008	.020	.268				-1							-.001	0
5	.218	.203	-.099	.030	-.151	.307					-1						.001	0
6	.044	.073	.071	.026	.056	.127						-1					-.002	0
7	.235	.076	-.019	-.006	.072	.140							-1				.001	0
8	.125	.223	-.079	.139	-.014	.204								-1			.001	0
9	.288	.046	.052	.073	.002	.202									-1		.001	0
10	.199	.095	-.001	.134	-.007	.160										-1	.002	0
a $1'D_{L_2L_2'}$	.489	.165	.032	.067	.034	.414	Σ	Check										
b $1'\Delta$	2.000	.720	2.400	4.140	2.550	5.600												
c $1'D_{L_2L_2'}\Delta$	.978	.119	.077	.277	.087	2.318	3.856	3.856										
d $1'(D_{L_2L_2'}\Delta)^{\frac{1}{2}}$	.989	.345	.277	.526	.295	1.522	3.954											
e $1'(D_{L_2L_2'}\Delta)(D_{L_2L_2'}\Delta)^{\frac{1}{2}}$	.978	.119	.077	.277	.087	2.316												
f $1'D_{b_3} = 1'(D_{L_2L_2'}\Delta)^{\frac{1}{2}}s_2$	17.758	6.195	4.974	9.445	5.297	27.329	70.999	70.998										
g $1'\Delta D_{b_3}^{-1}$	.113	.116	.482	.438	.481	.205	1.835											
h $1'D_R + 1'\Delta D_{b_3}^{-1}$	-1.033	1.036	1.402	1.258	1.311	1.065	7.105	7.105										

$$s_2 = \frac{T}{1'(D_{L_2L_2'}\Delta)^{\frac{1}{2}}1} = \frac{71}{3.954} = 17.956$$

9. Step 8 is repeated in rows *a* through *h* in Table 7, where a third approximation to the new test lengths is seen in row *f*.

Steps 5, 6, and 7 were again repeated to obtain a fourth approximation to the new test lengths. The values so obtained are shown, together with those of the preceding approximations, in Table 8. The vector of altered test lengths has not yet completely stabilized, but may be considered sufficiently so for all practical purposes.

10. The index of absolute prediction efficiency, λ, is computed as follows:

(*a*) To obtain the index λ, corresponding to the first approximation to optimal test lengths, calculate the product of each value in the L_1 matrix by the corresponding value in the validity matrix seen in Table 2, and sum the products. Thus the value of λ_1, seen as the first entry in the column to the far right in Table 8, is 2.053.

TABLE 8

Successive Approximations to $1'D_b$, for $T_1 = \frac{1}{2}T_0 = \frac{142}{2}$

Approx'n		1	2	3	4	5	6	Σ	Value of λ for successive values in L	
(.5) 1 D_a :	1	12.50	4.50	15.00	11.50	7.50	20.00	71.00	L_1	2.053
	2	16.96	5.96	6.25	10.33	6.12	25.36	70.98	L_2	2.086
	3	17.76	6.20	4.97	9.44	5.30	27.33	71.00	L_3	2.089
	4	17.90	6.30	4.61	9.10	4.81	28.28	71.00	L_4	2.090

(*b*) For λ_2, use the L_2 matrix instead of L_1, and repeat the procedure described in (*a*).

(*c*) To obtain any subsequent index, λ_i, substitute the L_i matrix for the L_1 matrix, while following the procedure indicated in (*a*).

The values of λ show only a very small increase in this particular illustration. From the initial value of 2.053 to a value of 2.090 corresponding to the fourth approximation, the increase is .037 or less than two per cent, although some of the test lengths are altered to a considerable extent. In the case of differential prediction (4), the increase in the efficiency index, ϕ, based on the same original data, was relatively larger, four per cent, for the fourth approximation to optimal test lengths.

Computations were also carried out with the original over-all testing time unchanged, and with the over-all testing time doubled. Three iterations were calculated for each of these conditions. The successive approximations to optimal test length, with the corresponding λ values, may be seen in Tables 9 and 10, respectively. Under these conditions, also, the increase in λ is small, although the alteration of some of the test lengths is relatively greater than that found in the first illustration. For unchanged over-all testing time, the increase in λ from the first to the third approximation is 1.5 per cent; for

TABLE 9

Successive Approximations to $1'D_b$, for $T_1 = T_0 = 142$

Approx'n		1	2	3	4	5	6	Σ	Value of λ for Successive Values in L	
(1.0) $1'D_a$:	1	25.00	9.00	30.00	23.00	15.00	40.00	142.00	L_1	2.203
	2	32.54	10.00	10.45	21.42	18.66	48.93	142.00	L_2	2.230
	3	32.87	9.70	8.21	20.21	21.61	49.40	142.00	L_3	2.234

TABLE 10

Successive Approximations to $1'D_b$, for $T_1 = 2T_0 = 2(142)$

Approx'n		1	2	3	4	5	6	Σ	Value of λ for Successive Values in L	
(2.0) $1'D_a$:	1	50.00	18.00	60.00	46.00	30.00	80.00	284.00	L_1	2.332
	2	63.21	15.81	16.23	43.98	56.11	88.66	284.00	L_2	2.373
	3	64.22	13.98	12.21	43.99	67.65	81.96	284.01	L_3	2.375

double the length of the original over-all testing time, the corresponding increase in λ is roughly two per cent. In the case of differential prediction (4), the corresponding increases in ϕ for three iterations were seven per cent and ten per cent, respectively.

It is obvious that for this particular example the prediction efficiency does not seem to be much improved by using optimal test lengths. Considerably more research may be required to indicate under what conditions, if any, prediction efficiency may be expected to vary appreciably as a function of variation in relative test length.

III. *Mathematical Derivation*

In (1) a procedure was developed for redistributing the specified over-all testing time for a battery of tests in such a way as to obtain optimal prediction of a single criterion variable. In this report, the procedure is extended to the problem of optimal prediction of multiple criteria. Let

$M \equiv$ the number of cases.
$n \equiv$ the number of predictors.
$N \equiv$ the number of criteria.
$Z \equiv$ an $(M \times n)$ matrix of test scores in a battery of altered length with the elements of Z of the form $(z_{ij} - \bar{z}_j)/(\sigma_{z_j}\sqrt{M})$.
$W \equiv$ an $(M \times N)$ matrix of criterion scores whose elements are deviate scores of the form $(w_{ij} - \bar{w}_j)/(\sigma_{w_j}\sqrt{M})$.
$B \equiv$ an $(n \times N)$ matrix of regression coefficients for estimating W from Z.
$r \equiv$ an $(n \times n)$ matrix of intercorrelations of tests of original lengths.
$\rho \equiv$ an $(n \times n)$ matrix of intercorrelations of tests of altered lengths.

$r_c \equiv$ an $(n \times N)$ matrix of validity coefficients for the tests of original lengths.
$\rho_c \equiv$ an $(n \times N)$ matrix of validity coefficients for the tests of altered lengths.
$D_a \equiv$ an $(n \times n)$ diagonal matrix of original test lengths.
$D_b \equiv$ an $(n \times n)$ diagonal matrix of altered test lengths.
$D_e = D_b D_a^{-1} \equiv$ the ratio of altered to original test lengths.
$D_{r_{ii}} \equiv$ the $(n \times n)$ diagonal matrix of reliability coefficients for the tests of original lengths.

The index of absolute prediction efficiency as defined in (3) is given by

$$\lambda = \operatorname{tr} C,$$

where C is the covariance matrix of predicted criterion scores in standard measure. Let

$$\delta = [I + (D_e - I)D_{r_{ii}}]D_e^{-1}. \tag{1}$$

Let

$$\epsilon = (ZB - W). \tag{2}$$

We wish to minimize the trace of $\epsilon'\epsilon$ with the constraining condition, $1'D_b1 = T$, where T represents the new over-all testing time and 1 is a column vector of all unit elements. From (2),

$$\epsilon'\epsilon = B'Z'ZB - B'Z'W - W'ZB + W'W. \tag{3}$$

From the definitions above,

$$Z'Z = \rho, \tag{4}$$

$$Z'W = \rho_c. \tag{5}$$

From (3), (4), (5)

$$\epsilon'\epsilon = B'\rho B - B'\rho_c - \rho_c'B + W'W. \tag{6}$$

Let

$$\psi = \operatorname{tr} \epsilon'\epsilon + \lambda 1'D_b1, \tag{7}$$

where λ is a Lagrangian multiplier. From (6) and (7)

$$\psi = \operatorname{tr}(B'\rho B - B'\rho_c - \rho_c'B) + N + \lambda 1'D_b1. \tag{8}$$

In (1) it is shown, for the case of a single criterion variable, that

$$\rho_c = \delta^{-1/2} r_c, \tag{9}$$

a relationship readily seen to hold for the case of multiple criteria. It was also shown in (1) that

$$\rho = \delta^{-1/2}(r - D_u + D_u D_a D_b^{-1})\delta^{-1/2}, \tag{10}$$

where D_u is defined as $I - D_{rii}$, a diagonal matrix of test unreliability coefficients. Substituting (9) and (10) in (8),

$$\psi = \operatorname{tr}\,[B'\delta^{-1/2}(r - D_u + D_u D_a D_b^{-1})\delta^{-1/2}B - B'\delta^{1/2} - r_c - r_c'\delta^{-1/2}B] + N + \lambda 1' D_b 1. \tag{11}$$

Let

$$B = \delta^{1/2}L, \tag{12}$$

$$r - D_u = R, \tag{13}$$

$$D_u D_a = \Delta. \tag{14}$$

Then (11) becomes

$$\psi = \operatorname{tr}\,[L'(R + \Delta D_b^{-1})L - L'r_c - r_c'L] + N + \lambda 1' D_b 1. \tag{15}$$

The unknowns on the right side of (15) are L, D_b, and λ. Differentiating (15) with respect to row vectors of L', and setting this derivative equal to zero,

$$\frac{\partial\psi}{\partial L'} = (R + \Delta D_b^{-1})L - r_c = 0,$$

or

$$(R + \Delta D_b^{-1})L = r_c\,. \tag{16}$$

Differentiating (15) with respect to D_b and setting this derivative equal to zero,

$$\frac{\partial\psi}{\partial D_b} = \lambda I - D_{LL'}\Delta D_b^{-2} = 0, \tag{17}$$

where $D_{LL'}$ is a diagonal matrix whose non-zero elements are the diagonal elements of LL'. From (17)

$$D_b = (D_{LL'}\Delta)^{1/2}/\lambda^{1/2}. \tag{18}$$

From (18)

$$1'D_b 1 = 1'(D_{LL'}\Delta)^{1/2}1/\lambda^{1/2}, \tag{19}$$

or, since we have the constraining condition, $1'D_b 1 = T$,

$$\lambda^{1/2} = 1'(D_{LL'}\Delta)^{1/2}1/T. \tag{20}$$

Substituting (20) in (18), we obtain

$$D_b = \frac{(D_{LL'}\Delta)^{1/2}T}{1'(D_{LL'}\Delta)^{1/2}1}. \tag{21}$$

From equations (16) and (21), the formulas are derived for solving for D_b by a series of successive approximations. From (16) we obtain

$$L = (R + \Delta D_b^{-1})^{-1} r_e \,. \tag{22}$$

Let

$$L_i = (R + \Delta D_{b_i}^{-1})^{-1} r_e \tag{23}$$

where

$$D_{b_1} = \frac{D_a T}{1' D_a 1}\,, \tag{24}$$

and

$$D_{b_{i+1}} = \frac{(D_{L_i L_i'} \Delta)^{1/2} T}{1' (D_{L_i L_i'} \Delta)^{1/2} 1}. \tag{25}$$

The first approximation to D_b is indicated by (24). The second and all subsequent approximations to D_b may be obtained by an iterative procedure based on (23) and (25). Thus, successive approximations to L_i and $D_{b_{i+1}}$ may be computed until D_b stabilizes satisfactorily.

The regression vectors for the tests of optimal length will be given by

$$B = \rho^{-1} \rho_e \,. \tag{26}$$

From (9), (10), (13), and (14)

$$B = \delta^{1/2} (R + \Delta D_b^{-1})^{-1} r_e \,, \tag{27}$$

and from (22) and (27),

$$B = \delta^{1/2} L, \tag{28}$$

where L has been stabilized through successive approximations.

Furthermore, it can be shown that the index of absolute prediction efficiency, λ, as defined in (3) is given by

$$\lambda = \operatorname{tr} L' r_e \,. \tag{29}$$

It may be that one or more elements of $D_{b_{i+1}}$, as given in (25), may approach zero as i increases. In this case it would be better to write (23) in the form

$$L_i = D_{b_i}^{1/2} (D_{b_i}^{1/2} R D_{b_i}^{1/2} + \Delta)^{-1} D_{b_i}^{1/2} r_e \,. \tag{30}$$

Although the computation of successive L_i matrices by means of (30) would be computationally more laborious than with (23), it would avoid difficulties resulting from one or more near vanishing elements of D_b.

The computational procedure presented in Section II is related to the mathematical derivation as follows:

Table 1 is based on equation (13).

Step 1 is based on equation (14).

Step 2 is given by equation (24).

Step 3 consists of calculating $D_{b_1}^{-1}$ from D_{b_1}.

Step 4 consists of calculating ΔD_b^{-1}.

Step 5 consists of calculating the matrix within the parentheses in equation (23) for the case $i = 1$.

Step 6 consists of the computation of the matrix L_1 from equation (23).

Step 7 consists of calculating D_{b_2} from equation (25) for the case $i = 1$.

Step 8 consists of the computation of the matrix L_2 from equation (23).

Step 9 consists of calculating D_{b_2} from equation (25).

In general, successive approximations to L and D_b are obtained by repeating steps 6 and 7 for successive values of i in equations (23) and (25).

Step 10 follows the procedure indicated by equation (29) to obtain successive values of λ.

REFERENCES

1. Horst, P. Determination of optimal test length to maximize the multiple correlation. *Psychometrika*, 1949, **14**, 79–88.
2. Horst, P. A technique for the development of a differential prediction battery. *Psychol. Monogr.*, 1954, **68**, No. 9 (Whole No. 380).
3. Horst, P. A technique for the development of a multiple absolute prediction battery. *Psychol. Monogr.*, 1955, **69**, No. 5 (Whole No. 390).
4. Horst, P. Optimal test length for maximum differential prediction. *Psychometrika*, 1956, **21**, 51–66.
5. Horst, P. Servant of the human sciences, Chapt. 21, Section 7. Unpublished manuscript. Division of Counseling and Testing Services, University of Washington.

Manuscript received 3/2/55

Revised manuscript received 4/28/55

PSYCHOMETRIKA—VOL. 21, NO. 2
JUNE, 1956

A LEAST SQUARES SOLUTION FOR PAIRED COMPARISONS WITH INCOMPLETE DATA*†

HAROLD GULLIKSEN

PRINCETON UNIVERSITY

AND

EDUCATIONAL TESTING SERVICE

A precise and rapid procedure has been devised for dealing with a matrix of incomplete data in paired comparisons. This method should increase the general applicability of paired comparisons since experiments involving large numbers of stimuli may now be shortened to feasible experimental proportions. Also, we may now use sets of stimuli which cover a wide range, resulting in a considerable number of 100 per cent vs. 0 per cent judgments, and still give a precise solution depending equally on each of the observations.

In scaling by paired comparisons, many cases arise where one does not have complete data. In any problem where the range of the set of stimuli is great in relation to the discriminal dispersion, there will be no usable data for the extreme comparisons. This is usually the case in dealing with construction of scales for various sensory areas, such as brightness, hue, pitch, or loudness. In other cases the experiment is made less laborious for the subject by not requiring all comparisons. If one is interested in studies of value judgments (6) where composite objects are used, such as (a and b vs. c) or (a and b vs. c and d), then it is also highly desirable not to include all possible comparisons. For example, it may be well to omit comparison (a and b vs. b) or the comparison (a and b vs. a and c). Such situations might arise in studying preferences for various foods, gifts, individuals, activities, or goals. Any of the types of situations indicated above may give rise to a matrix of incomplete data for which a reasonably precise solution is desired. A solution for paired comparisons with incomplete data for the case in which the correlations are equal and the ratios of the discriminal dispersions are known is presented here. The general usability of paired comparisons, especially in fields such as sensation and value judgments, will be

*This study was supported in part by the Office of Naval Research Contract N6onr-270-20 with Princeton University and by the National Science Foundation Grant G-642. The opinions expressed are, of course, those of the author and do not represent attitudes or policies of the Office of Naval Research or of the National Science Foundation.

†The author wishes to acknowledge helpful suggestions and comments received from Frederic M. Lord, Warren S. Torgerson, and Ledyard R Tucker. Thanks are also due to Mrs. Gertrude Diederich for some of the tabulating and computing for the illustrative problem.

greatly enhanced by a precise method for utilizing all the available data when dealing with an incomplete matrix.

The procedure to be presented follows the general method for handling incomplete data outlined by Horst (2, pp. 419–430), Kempthorne (3), or Kendall (4). The least squares solution for paired comparisons, when correlations are equal and ratios of discriminal dispersions are known, has been indicated by Mosteller (5). The problem may be stated in the following manner for the case in which complete data are available:

For the case of errorless data let S_i^*, S_j^* $(i = 1, \cdots, n; j = 1, \cdots, n)$ be the scale values of the stimuli. Define the distance $D_{ij}^* \equiv S_i^* - S_j^*$. From the values of p_{ij} (the experimental proportions of judgments $i > j$) we use a normal curve or some other assumption to determine values of D_{ij}, which are taken as estimates of D_{ij}^*.

The scaling problem is to determine values S_i, S_j $(i = 1, \cdots, n; j = 1, \cdots, n)$ such that $\sum_{i=1}^{n} \sum_{j=1}^{n} (D_{ij} - S_i + S_j)^2$ is a minimum. The subscripts i and j are alternative subscripts each designating the stimuli from 1 to n.

Least Squares Solution for Incomplete Data

For the case in which only incomplete data are available we may write

$$Q = \sum_{i,j}^{n_{ij}} (D_{ij} - S_i + S_j)^2, \tag{1}$$

using n_{ij} to indicate that the summation is over the available data. The scaling problem is to determine S_i and S_j $(i, j = 1, \cdots, n)$ such that Q is a minimum. It should be noted that since the matrix of D_{ij} for complete data is skew symmetric the matrix of D_{ij} for incomplete data will also be skew symmetric.

To determine the unknown S values that will make Q a minimum, take the partial derivative of Q with respect to S_i, giving

$$\frac{1}{2}\frac{\partial Q}{\partial S_i} = -\sum_{j}^{n_{ij}} (D_{ij} - S_i + S_j) + \sum_{j}^{n_{ij}} (-D_{ij} + S_i - S_j). \tag{2}$$

One term represents the partial derivative for the *row* in which S_i occurs in each cell, and the other term represents the partial derivative for the *column* of entries each of which contains S_i. The first term is identical with the second so we may write

$$\frac{1}{2}\frac{\partial Q}{\partial S_i} = 2 \sum_{j}^{n_{ij}} (S_i - S_j - D_{ij}). \tag{3}$$

If n_i represents the number of observations in row (or column) i, then simplifying and setting the partial derivative equal to zero gives

$$n_i S_i - \sum_{j}^{n_i} S_j - \sum_{j}^{n_i} D_{ij} = 0 \qquad (i = 1, \cdots, n). \tag{4}$$

This set of n equations may be expressed in matrix notation. Let

$S \equiv$ the column vector of elements S_i.

$1 \equiv$ a column vector of 1's.

$D \equiv$ the matrix with elements D_{ij} wherever observations occur, zero for each missing observation, and zeroes in the principal diagonal. An illustrative D matrix is shown in Table 3 provided a zero is put in for each missing entry.

$M \equiv$ a matrix constructed from D according to the following rules: Enter -1 in M for each cell entry in D where data exist. Enter 0 in all other off-diagonal cells. The entry in the diagonal cells in n_i—the number of observations in row (or column) i. Note that M is a symmetric matrix and that the sum of each row (and column) is zero, and hence M^{-1} does not exist.

$Z \equiv D1$, a column vector, the sum of the rows of matrix D.

Using this notation, the set of equations (4) may be written

$$MS = Z. \tag{5}$$

Matrix M has no inverse, but, in general, its first minor will have an inverse. If we specify an origin by some method such as arbitrarily setting the first element of S equal to zero, a solution for the remaining elements of S is given by deleting the first element from Z and S, and the first row and column from M, giving

$$M_{11}S_1 = Z_1 , \tag{6}$$

which has the solution

$$S_1 = M_{11}^{-1}Z_1 . \tag{7}$$

It may be noted that for complete data M is an $n \times n$ matrix with $n - 1$ in the principal diagonal and -1 in each off-diagonal cell. The inverse of M_{11} for the complete data case is a matrix with $2/n$ in each principal diagonal cell and $1/n$ in each off-diagonal cell. In this case the solution given in (7) corresponds to the solution given by Mosteller (5, equation 10).

Subsequent computations and the present analysis will be facilitated by the use of two other matrices:

$N \equiv$ a diagonal matrix with reciprocal of $(n_i + 1)$ as the element in the ith diagonal cell and zero in each off-diagonal cell.

$L \equiv$ a matrix constructed from M by putting a zero wherever there is a zero in M and $+1$ in all other cells.

Note that

$$M = N^{-1} - L. \tag{8}$$

Using this notation (5) becomes

$$Z = N^{-1}S - LS. \tag{9}$$

For the matrix of complete data L becomes a matrix with unity in every cell and thus LS is a vector of constants. The origin of S can be chosen so that $LS = 0$, whereupon (9) can be solved giving

$$S = NZ, \tag{10}$$

which is also equivalent to the solution given by Mosteller (5, equation 10).

Iterative Solution

However, for the case of incomplete data where M_{11}^{-1} is difficult to compute, a solution may be found quickly by an iterative procedure. The procedure given here is an analytical analog of a graphical iterative procedure that was devised by Mrs. Gertrude Diederich. It was suggested by the procedure outlined by Garner and Hake (1). It corresponds essentially to taking their solution as a *first approximation* and then correcting it to obtain successively closer approximations.

The iterative procedure proposed here for paired comparison is outlined below. We begin by taking a trial set of scale values. Since this corresponds to *assuming* a set of values for S we may say that T_1 designates this first set of trial values. The discrepancies between the *predicted values* (MT_1) and the *experimental values* (Z) are found by taking differences $(Z - MT_1) \equiv ME_1$. The average of the discrepancies for each scale value is then used to correct that value. This corresponds to taking the correction NME_1 and computing the second set of trial values by setting $T_2 = T_1 + NME_1$.

Graphically this correction corresponds to changing each element of T_1 by the average of all the discrepancies found for that element. When tried graphically, such a procedure gave convergence as far as detectable from a reasonably large graph in two or three trials. Although no analytical proof of convergence has yet been found, it seems intuitively reasonable that utilizing the same correction analytically would give a reasonable approximation to the solution. In any case, the discrepancies (ME_i) would be computed at each step, so that failure of convergence could be readily detected. This process of computing $T_{i+1} = T_i + NME_i$ is continued until the discrepancies and the correction terms are negligible.

A little matrix substitution shows that if $T_3 = T_2 + NME_2$, then $T_3 = T_1 + N(I + LN)ME_1$. Thus, since the "two step" correction is very easy to compute, we find that $T_2 = T_1 + N(I + LN)ME_1$. In the computational procedure it is also necessary to adjust the *general scale* (variance) of the trial values to agree with that of the observations. Generalizing these computations we have the following steps for an iterative computation of scale values:

1. Select T_1, any set of trial S values. The values $0, 1, 2, 3, 4, \cdots, (n-1)$ will suffice although convergence will probably be slightly more rapid if the average difference of paired D_{ij}-values is used.

2. Compute $MT_{1'}$.
3. Adjust the general scale of $T_{1'}$ to Z by computing

$$a^2 = Z'Z/(T_{1'}'M'MT_{1'}).$$

4. Set $T_1 = aT_{1'}$, and $MT_1 = aMT_{1'}$.
5. Compute the error of approximation

$$ME_1 = Z - MT_1.$$

6. Compute the correction

$$C_1 = N(I + LN)ME_1 = N(I + I - MN)ME_1 .$$

7. $$T_{2'} = T_1 + C_1 + b_1,$$

where b is an additive constant to adjust the origin to some convenient value.

Repeat beginning with Step 1 until the error of approximation obtained in Step 5 is negligible.

As a guide to computation it should be noted that C is computed most readily by the following sequence: multiplying to obtain NME_1; selective addition of elements in NME_1 yields $LNME_1$; another addition of ME_1 to $LNME_1$ gives $(I + LN)ME_1$; multiplying gives $N(I + LN)ME_1$. In two problems worked by this procedure each element of the vector indicated in Step 5 was reduced to .005 or less on the third approximation. One problem is presented here to illustrate the results of this procedure.

Illustrative Problem

A food preference questionnaire was constructed using five different main courses and the ten composites formed by taking all possible pairs of these five. For each choice of the form (i vs. j) or (i and j vs. g) or (i and j vs. g and h) the subject was asked to indicate his preference. Three sample choices are shown in Table 1 together with the code used in Tables 2, 3, and 4 for each of the five foods.

This questionnaire was given to 92 college students with the result shown in Table 2. The number in each cell indicates number of votes for the item indicated at the beginning of the row when it was paired with the item at the top of the column. For example, when given the choice between Tongue and Pork, 68 persons chose Pork while 24 chose Tongue. Four of the paired comparisons were made by only 91 persons. These were utilized rather than lowering the number of complete cases to 88.

Comparisons of the form (i and j vs. i) were omitted from the questionnaire. As the results turned out, it would have been interesting to have had such items. A few comparisons of the form (i and j vs. i and k) were included. These, however, were eliminated in the analysis since it was not clear from the results whether the subjects were judging one *composite* against another or were merely ignoring the common element and comparing (j vs. k) as a (1 vs. 1) comparison. The item (P and S vs. T and L) was given in the middle

TABLE 1

Sample Questionnaire Items

		Code used in Tables 2, 3, and 4
38.	☐ Roast Rib of Prime Beef	(B)
	☐ Roast Loin of Pork	(P)
39.	☐ Roast Rib of Prime Beef Sirloin Steak	(B) (S)
	☐ Boiled Smoked Beef Tongue Loin Lamb Chop	(T) (L)
40.	☐ Sirloin Steak	(S)
	☐ Roast Loin of Pork Boiled Smoked Beef Tongue	(P) (T)

TABLE 2

Food Preference Data*

	TP	T	TL	P	TB	PL	L	TS	PB	B	PS	LB	S	LS	BS
TP	X	-	-	-	-	-	18	-	-	2	-	5	2	4	0
T	-	X	-	24	-	16	13	-	8	1	9	3	0	1	1
TL	-	-	X	37	-	-	-	-	13	5	10 13	-	2	-	3
P	-	68	54	X	39	-	21	30	-	4	-	3	0	6	5
TB	-	-	-	53	X	37	40	-	-	-	17	-	16	9	-
PL	-	76	-	-	55	X	-	45	-	22	-	-	12	-	2
L	73	79	-	71	52	-	X	46	31	13	22	-	6	-	10
TS	-	-	-	62	-	47	46	X	31	32	-	24	-	-	-
PB	-	84	78	-	-	-	60	61	X	-	-	-	21	15	-
B	90	91	87	88	-	70	79	60	-	X	35	-	19	23	-
PS	-	83	82 79	-	75	-	70	-	-	57	X	43	-	-	-
LB	87	89	-	89	-	-	-	68	-	-	49	X	37	-	-
S	90	92	90	92	76	80	86	-	71	73	-	55	X	-	-
LS	88	91	-	86	83	-	-	-	77	69	-	-	-	X	-
BS	92	91	89	87	-	90	82	-	-	-	-	-	-	-	X

*Number of votes for stimulus listed at side when it was paired with the stimulus listed at the top. The key for stimulus abbreviations is given in Table 1.

of the questionnaire and again near the end with the results shown in Table 2. In converting this cell to a "distance" entry the average of the two different distances was used.

The values shown in Table 2 were converted to normal deviate values by use of a table of the normal curve. An adjustment for different discriminal dispersions was made on the basis of a priori considerations. Since the number of items involved in the three types of judgments [the (1 vs. 1), the (2 vs. 1), and the (2 vs. 2)] were in the ratio of 2 to 3 to 4, the *variances*, or the squares of the discriminal dispersions, might reasonably be assumed to have the same ratio. To approximate this ratio the entries from the normal table were used directly for all comparisons of the (2 vs. 1) and (1 vs. 2) type. For comparisons of the (2 vs. 2) type these values were multipled by 1.2 (which equals approximately $\sqrt{4}/\sqrt{3}$). For comparisons of the (1 vs. 1) type the normal curve values were multiplied by 0.8 (which equals approximately $\sqrt{2}/\sqrt{3}$).

Also, the three types of judgments were scaled separately and the scales seemed to agree reasonably well after the adjustment for the assumed differences in discriminal dispersions were made.

Judgments of the form 92 vs. 0 were converted into "lower bound" values by assigning the normal deviate for 91.5 vs. 0.5. As the results turned out, these values were not systematically lower than predicted values so perhaps such an approximation is possible where only a few 100 per cent judgments are found.

Table 3 gives the resulting D matrix, provided a zero is substituted for each missing entry. The sums of the rows shown in Table 3 are elements of the column vector Z.

The results of the computations indicated in Steps 1 to 7 are shown in Table 4. Since a in Step 3 turned out to be nearly unity each time, only the trial vectors T_1, T_2, and T_3 are shown while $T_{1'}$, $T_{2'}$, and $T_{3'}$ are not shown. The vectors indicating the error of approximation to a least squares fit, ME_1, ME_2, and ME_3 are shown, and indicate extremely rapid convergence.

The correction vectors $C_1 + b_1$ and $C_2 + b_2$ indicated in Step 7 are also shown. It should be noted that each b_i has been chosen so that the element corresponding to "Tongue and Pork" (TP) remains at zero. Some such adjustment facilitates comparisons of the successive approximations and does not affect goodness of fit since the origin is arbitrary.

A very interesting regularity appears in these results. The components in order from least preferred to most preferred are Tongue, Pork, Lamb, Beef, and Steak. Adding Lamb, Beef, or Steak *increases* the value of a composite; while adding Tongue or Pork *decreases* the value of the composite. Thus, the evidence, purely from the general ordering of the stimuli, places Tongue and Pork as negative and Lamb, Beef, and Steak as positive values. A more precise determination of the zero point will be given in a later article.

TABLE 3

Food Preferences

Experimental Inter-Stimulus Distances

(The key for stimulus abbreviations is given in Table 1)

	TP	T	TL	P	TB	PL	L	TS	PB	B	PS	LB	S	LS	BS	Sums
TP	0	X	X	X	X	X	-0.850	X	X	-2.020	X	-1.920	-2.020	-2.052	-3.060	-11.922
T	X	0	X	-0.512	X	-0.940	-0.856	X	-1.360	-1.832	-1.290	-1.840	-2.040	-2.290	-2.290	-15.250
TL	X	X	0	-0.240	X	X	X	X	-1.270	-1.600	-1.380	X	-2.020	X	-2.208	-8.718
P	X	0.512	0.240	0	-0.190	X	-0.592	-0.450	X	-1.368	X	-1.840	-2.040	-1.510	-1.600	-8.838
TB	X	X	X	0.190	0	-0.300	-0.160	X	X	X	-1.080	X	-0.940	-1.548	X	-3.838
PL	X	0.940	X	X	0.300	0	X	-0.036	X	-0.710	X	X	-1.120	X	-2.424	-3.050
L	0.850	0.856	X	0.592	0.160	X	0	0.000	-0.410	-0.856	-0.710	X	-1.208	X	-1.230	-1.956
TS	X	X	X	0.450	X	0.036	0.000	0	-0.504	-0.390	X	-0.768	X	X	X	-1.176
PB	X	1.360	1.270	X	X	X	0.410	0.504	0	X	X	X	-0.740	-1.176	X	1.628
B	2.020	1.832	1.600	1.368	X	0.710	0.856	0.390	X	0	-0.300	X	-0.656	-0.670	X	7.150
PS	X	1.290	1.380	X	1.080	X	0.710	X	X	0.300	0	-0.096	X	X	X	4.664
LB	1.920	1.840	X	1.840	X	X	X	0.768	X	X	0.096	0	-0.250	X	X	6.214
S	2.020	2.040	2.020	2.040	0.940	1.120	1.208	X	0.740	0.656	X	0.250	0	X	X	13.034
LS	2.052	2.290	X	1.510	1.548	X	X	X	1.176	0.670	X	X	X	0	X	9.246
BS	3.060	2.290	2.208	1.600	X	2.424	1.230	X	X	X	X	X	X	X	0	12.812

TABLE 4

Iterative Procedure for Determining Scale Values

	$Z - MT_1$	$N(I + LN)ME_1$		$T_1 + C_1$	$Z - MT_2$	$N(I + LN)ME_2$	$S \cong T_3$	$Z - MT_3$
Stimuli[a]	T_1[b]	ME_1[c]	C_1[d]	T_2[b]	ME_2[c]	C_2[d]	T_3[b]	ME_3[c]
TP	0.0	.778	.00000	.000	.049	.00000	.000	.003
T	0.3	-.350	-.15379	.146	-.010	-.00858	.137	.002
TL	0.4	.082	-.12091	.279	-.004	-.00945	.270	-.004
P	0.7	-.138	-.15236	.548	.020	-.00699	.541	.002
TB	1.0	-.338	-.15929	.841	-.031	-.01126	.830	-.005
PL	1.0	.550	-.06575	.934	.027	-.00594	.928	.002
L	1.2	-.156	-.14974	1.050	.020	-.00681	1.043	.003
TS	1.3	-.476	-.19896	1.101	-.035	-.01331	1.088	-.005
PB	1.6	-.072	-.14123	1.459	-.017	-.01087	1.448	-.001
B	1.9	-.350	-.14431	1.756	-.031	-.00986	1.746	-.002
PS	1.9	.164	-.11220	1.788	.008	-.00760	1.780	.003
LB	2.1	.114	-.10014	2.000	.002	-.00709	1.993	-.001
S	2.3	.234	-.09550	2.205	-.003	-.00793	2.197	.000
LS	2.4	.346	-.07185	2.328	.028	-.00364	2.324	.004
BS	2.8	-.388	-.16790	2.362	-.023	-.00986	2.622	-.001

[a] Stimuli abbreviations are defined in Table 1.

[b] T_1, T_2, and T_3 are successive trial values for S (scale values of the stimuli)

[c] ME_1, ME_2, and ME_3 are the errors of approximation.

[d] C_1 and C_2 are the correction terms.

It turned out to be necessary to carry the computations of Step 6 to five decimals in order to be certain of having the new trial value for T correct to three decimal places. Amount of lowest entry has been subtracted from each element in C so that correction on lowest term is zero. This does not affect fit but merely keeps origin of T_1 constant.

REFERENCES

1. Garner, W. R., and Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, **58**, 1951, 446–459.
2. Horst, P. The prediction of personal adjustment. *Social Science Research Council Bulletin No. 48*, 1941.
3. Kempthorne, O. The design and analysis of experiments. New York: Wiley, 1952, pp. 96–98.
4. Kendall, M. G. The advanced theory of statistics, Vol. II. London: Griffin, 1948, pp. 229–230.
5. Mosteller, F. Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, **16**, 1951, 3–9.
6. Thurstone, L. L. The zero point in scaling. (Unpublished study.)

Manuscript received 11/29/54

Revised manuscript received 4/29/55

PSYCHOMETRIKA—VOL. 21, NO. 2
JUNE, 1956

PROPORTIONAL PROFILES AND LATENT STRUCTURE

W. A. GIBSON

PERSONNEL RESEARCH BRANCH
DEPARTMENT OF THE ARMY*

The identity of problem and solution in Lazarsfeld's latent structure analysis and Cattell's proportional profiles is pointed out. Anderson's latent structure solution is adapted to proportional profiles to yield a possible solution for the communality and rotational problems in factor analysis. A numerical example of the latter is provided.

A Formal Identity

Two recent articles (4, 6) have presented identical solutions to the same geometric problem which has appeared in two different contexts. The two different contexts are Lazarsfeld's new latent structure model (7) and Cattell's concept of proportional profiles in factor analysis (3). The common problem may be stated as follows:

> Given two Gramian matrices, R_1 and R_2, of order n and rank m, to find corresponding orthogonal factor matrices, V_1 and V_2, that are proportional by columns and are known to exist from theoretical considerations.

Let F_1 be an arbitrary orthogonal factorization of R_1, and let Λ be the orthogonal transformation from F_1 to V_1. Also let D be the diagonal matrix of proportionality constants. Then

$$R_1 = F_1F_1' = V_1V_1', \tag{1}$$

$$V_1 = F_1\Lambda, \tag{2}$$

$$V_2 = V_1D = F_1\Lambda D, \tag{3}$$

and

$$R_2 = V_2V_2' = V_1D^2V_1' = F_1\Lambda D^2\Lambda'F_1'. \tag{4}$$

The solution is to form the matrix

$$\begin{aligned} A &= (F_1'F_1)^{-1}F_1'R_2F_1(F_1'F_1)^{-1} \\ &= (F_1'F_1)^{-1}F_1'F_1\Lambda D^2\Lambda'F_1'F_1(F_1'F_1)^{-1} \\ &= \Lambda D^2\Lambda', \end{aligned} \tag{5}$$

*This paper was initiated at the University of North Carolina and completed at the Center for Advanced Study in the Behavioral Sciences.

factor it by any principal components method, and normalize the resulting factor matrix by columns to obtain Λ. Then V_1 and V_2 are given by (2) and (3) in turn, D having been determined in the normalizing process.

Since the diagonal elements in D^2 are the characteristic roots of A, this solution cannot yield a unique Λ unless all diagonal elements in D^2 (hence also in D) are different. This creates no serious difficulties in latent structure analysis, for there the empirical data can usually be combined in such a way as to make all diagonal values in D^2 quite distinct from each other. In proportional profiles, however, there is no such manipulatory freedom. There R_1 and R_2 are correlation matrices for the same tests based on two different samples, and V_1 and V_2 are corresponding factor matrices. The diagonal values in D indicate differential degrees of selection on the factors in samples 1 and 2. In practice it may often be difficult to find two samples in which the degrees of selection on the various factors are all quite different. This will create problems of slow convergence or even indeterminacy for some of the factors. While such problems will not be discussed here, their possible seriousness should not be minimized in evaluating the notion of proportional profiles.

It is instructive to consider the special case where F_1 is the principal components analysis of R_1. The matrix $F_1'F_1$ then is diagonal and contains the characteristic roots of R_1 in its diagonal cells. The matrix $(F_1'F_1)^{-1}$ is also diagonal, with the reciprocals of the characteristic roots of R_1 as its diagonal entries. The pre- and post-multiplication of the matrix $F_1'R_2F_1$ by this diagonal matrix in (5) produces a situation which is exactly the opposite of what might be expected. The last diagonal values in $(F_1'F_1)^{-1}$ are the largest, so that the last vectors in A are likely to be the longest. Hence the first principal components in A are determined largely by the last principal components in R_1. Since the last components in R_1 are probably most seriously affected by error, less confidence can be placed in the first columns of V_1 and V_2 than in the last. This line of reasoning applies also to the centroid F_1 insofar as it approximates the principal components F_1.

Anderson's Latent Structure Solution and Proportional Profiles

Since the geometric problem in latent structure analysis and in proportional profiles is the same, it follows that any solution for the latent structure equations is also a solution for proportional profiles. A recent solution for the former, developed by Anderson (1) and extended by Gibson (5), has the advantage over Green's solution of avoiding the estimation of any unknown elements in the manifest matrices (such as the missing diagonal terms). It is not difficult to adapt this solution to proportional profiles so as to eliminate both the communality and rotational problems in factor analysis, if Cattell's concept is accepted and is workable.

Let the two correlation matrices, R_1 and R_2, be rearranged, if need be,

so that the two sub-matrices of order $n - m$ by m in the lower left corner of R_1 and R_2 have rank m. Let these two sub-matrices be designated P_1 and P_2, respectively. Let the subscript a refer to the last $n - m$ test variables, and let b stand for the first m test variables. Then

$$P_1 = V_{a1}V'_{b1}, \tag{6}$$

and

$$P_2 = V_{a2}V'_{b2} = V_{a1}D^2V'_{b1}. \tag{7}$$

That is, V_{a1} and V_{a2} are made up of the last $n - m$ rows of V_1 and V_2, respectively, and V'_{b1} and V'_{b2} are square matrices consisting of the first m columns of V'_1 and V'_2, respectively.

Now form the matrix

$$\begin{aligned} B &= (P'_1P_1)^{-1}P'_1P_2 \\ &= (V_{b1}V'_{a1}V_{a1}V'_{b1})^{-1}V_{b1}V'_{a1}V_{a1}D^2V'_{b1} \\ &= V'^{-1}_{b1}(V'_{a1}V_{a1})^{-1}V^{-1}_{b1}V_{b1}V'_{a1}V_{a1}D^2V'_{b1} \\ &= V'^{-1}_{b1}D^2V'_{b1}. \end{aligned} \tag{8}$$

The next task is to obtain the characteristic roots and a set of right-sided characteristic vectors of B, for it turns out that the roots are the diagonal entries in D^2 and the vectors are the columns of $V'^{-1}_{b1}K$, K being an arbitrary unknown diagonal. Thus

$$\begin{aligned} BV'^{-1}_{b1}K &= V'^{-1}_{b1}D^2V'_{b1}V'^{-1}_{b1}K \\ &= V'^{-1}_{b1}D^2K = V'^{-1}_{b1}KD^2. \end{aligned} \tag{9}$$

Once a matrix $V'^{-1}_{b1}K$ is obtained, (6) can be post-multiplied by it to give

$$C = P_1V'^{-1}_{b1}K = V_{a1}V'_{b1}V'^{-1}_{b1}K = V_{a1}K. \tag{10}$$

Thus V_{a1} becomes available except for unknown multipliers on its columns. It happens that these multipliers are quite readily obtained in latent structure analysis because of the way in which the empirically given matrices are bordered. It is necessary to adopt a different approach in proportional profiles. Let the symmetric sub-matrix of order $n - m$ in the lower right corner of R_1 be designated Q_1. Then

$$Q_1 = V_{a1}V'_{a1} = (V_{a1}K)K^{-2}(KV'_{a1}) = CK^{-2}C'. \tag{11}$$

Except for its diagonal terms, the matrix Q_1 is entirely given, as is the matrix C. Assume for the moment that the diagonal terms in Q_1 are given, and define

$$G = C(C'C)^{-1}. \tag{12}$$

Then K^{-2} can be obtained from (11) as follows:

$$K^{-2} = (C'C)^{-1}C'Q_1C(C'C)^{-1} = G'Q_1G. \tag{13}$$

Given K^{-2}, K^{-1} is easily formed, after which V_{a1} can be determined by rewriting (10) in the form

$$V_{a1} = CK^{-1}. \tag{14}$$

This suggests a simple iterative procedure in which V_{a1} is first approximated by inserting rough diagonal estimates into Q_1 and applying (13) and (14), then improved diagonal estimates obtained from the first V_{a1} permit a second cycle of the same kind, and so on until further iteration makes no important change in V_{a1} from one cycle to the next. For large Q_1 matrices, little or no iterating will be needed, since the products involving the estimated diagonals will constitute such a small part of the sums of products. With small Q_1 matrices the iterations will not be very time-consuming. To save time in iterating it may be worth while to express Q_1 as the sum of two matrices, one of them being a diagonal matrix containing the unknown diagonal elements of Q_1, and the other being Q_1 with zero diagonal elements. Designate these two matrices E and Q_{10}, respectively. Then

$$K^{-2} = G'(Q_{10} + E)G = G'Q_{10}G + G'EG. \tag{15}$$

Only the last term in (15) changes from one trial to the next, and it is rapidly formed. The matrix G remains constant throughout the iterative process.

Given V_{a1}, V_{b1} can be solved for from (6), (10), and (12) as follows:

$$\begin{aligned} V_{b1} &= P_1'V_{a1}(V_{a1}'V_{a1})^{-1} \\ &= P_1'V_{a1}K(KV_{a1}'V_{a1}K)^{-1}K \\ &= P_1'C(C'C)^{-1}K = P_1'GK. \end{aligned} \tag{16}$$

The last member of (16) has been adjusted so as to involve the same matrix G that is used in the iteration to determine V_{a1}. Thus G is made to serve two purposes.

Once V_{a1} and V_{b1} are known, all that remains to be done is to form V_1 from them to give the "true" factorization of R_1, and then to compute V_2 from the first part of (3), the matrix D having been formed from the square roots of the characteristic roots of B. The goodness of fit of the two factorizations V_1 and V_2 is indicated by the agreement between the first and third members of (1) and between the first and second members of (4).

It is to be noted that this adaptation of Anderson's solution to the problem of proportional profiles assumes the number of factors to be known at the start. This is not a serious drawback, however, for there are many ways of estimating the rank of a correlation matrix, and the inverting of $P_1'P_1$ can

be done in such a way (cf. **8**, pp. 46-48) that a change in its order can be accomplished without serious loss of computing time.

Several equations in this paper have least squares properties that should perhaps be listed explicitly. The first is (5), which is the best fitting solution for $\Lambda D^2 \Lambda'$ in (4). The next is (8), which is the least squares solution of B in the equation $P_2 = P_1 B$. Then (13) is the best fitting solution for K^{-2} in (11). Finally, (16) is the least squares solution for V_{b1} in (6).

A Fictitious Example

As an example of the present solution for proportional profiles, consider the two fictitious correlation matrices, R_1 and R_2, that are shown in Tables 1 and 2. They were generated from a simple structure V_1 and V_2 that originally were strictly proportional by columns, but that subsequently were "blurred" by adding small random increments to each of their entries. Thus there exists no perfect proportional profiles fit for R_1 and R_2, so that the various least squares properties of the solution will have definite advantages in this example.

Inspection of R_1 and R_2 suggests that three factors will probably account for both, and that the 7×3 sub-matrix in the lower left corner of each probably has rank 3. These two sub-matrices are therefore designated P_1 and P_2, respectively, and (8) is used to form from them the matrix B that is shown in Table 3.

Table 4 is the matrix D^2, containing in its diagonal cells the characteristic roots of B. These characteristic roots are obtained by forming and solving the characteristic equation of B(cf. **8**, pp. 26-28 and 44-45). The algebraic sign of the characteristic roots in Thurstone's discussion is the opposite of the convention being used here. The three columns of Table 5 are a set of right-sided characteristic vectors of B. Column A, for example, is a solution to the set of homogeneous linear equations whose coefficients matrix is the matrix B with its first characteristic root subtracted from each of its diagonal cells (cf. **2**, pp. 250-251). Column B is the same thing using the second characteristic root, etc. Each such set of homogeneous linear equations is readily converted into a consistent set of m non-homogeneous linear equations in $m - 1$ unknowns by arbitrarily fixing one of the unknowns at some convenient value. Here the first element of the first characteristic vector was set equal to unity. The resulting over-determined system can be solved by any method. Here it was done by least squares because of the slight inconsistency resulting from rounding.

The next step is to form the matrix C from P_1 and $V_{b1}'^{-1}K$ by means of (10). C is shown in Table 6; (10) indicates that it is proportional by columns to V_{a1}.

The task now becomes one of extracting V_{a1} from C by determining K^{-1}. As a first step the matrix G is formed by means of (12); G is shown in

TABLE 1
A Fictitious R_1

	1	2	3	4	5	6	7	8	9	10
1		.05	.02	.25	.02	.02	.18	.13	.06	.18
2	.05		.10	.07	.46	.08	.36	.11	.24	.31
3	.02	.10		.00	.14	.39	.10	.35	.33	.20
4	.25	.07	.00		.05	.00	.19	.11	.05	.18
5	.02	.46	.14	.05		.12	.39	.14	.28	.34
6	.02	.08	.39	.00	.12		.08	.38	.36	.20
7	.18	.36	.10	.19	.39	.08		.17	.23	.35
8	.13	.11	.35	.11	.14	.38	.17		.35	.27
9	.06	.24	.33	.05	.28	.36	.23	.35		.30
10	.18	.31	.20	.18	.34	.20	.35	.27	.30	
		P_1					Q_1			

TABLE 2
A Fictitious R_2

	1	2	3	4	5	6	7	8	9	10
1		.13	.08	.46	.08	.06	.32	.23	.13	.37
2	.13		.05	.16	.50	.01	.45	.10	.25	.37
3	.08	.05		.08	.10	.20	.07	.24	.20	.15
4	.46	.16	.08		.10	.05	.37	.26	.14	.42
5	.08	.50	.10	.10		.05	.40	.12	.27	.34
6	.06	.01	.20	.05	.05		.02	.20	.16	.10
7	.32	.45	.07	.37	.40	.02		.19	.24	.47
8	.23	.10	.24	.26	.12	.20	.19		.23	.28
9	.13	.25	.20	.14	.27	.16	.24	.23		.28
10	.37	.37	.15	.42	.34	.10	.47	.28	.28	
		P_2								

TABLE 3
$B=(P_1'P_1)^{-1}P_1'P_2$

	1	2	3
1	1.76	.29	.25
2	.07	1.15	.00
3	.04	.18	.55

TABLE 4
D^2

	A	B	C
A	1.80		
B		1.13	
C			.54

TABLE 5
$V_{b1}'^{-1}K$

	A	B	C
1	1.00	-.33	-.21
2	.11	1.00	.02
3	.02	-.33	1.00

TABLE 6
$C=P_1V_{b1}'^{-1}K$

	A	B	C
4	.26	-.01	-.05
5	.07	.41	.14
6	.04	-.06	.39
7	.22	.27	.07
8	.15	-.05	.32
9	.09	.11	.32
10	.22	.18	.17

TABLE 7
$G=C(C'C)^{-1}$

	A	B	C
4	2.35	-.84	-.72
5	-.97	1.76	.24
6	-.32	-.47	1.17
7	.89	.63	-.29
8	.74	-.75	.71
9	-.26	.21	.82
10	.95	.18	.05

Table 7. Now the constant part ($G'Q_{10}G$) of (15) can be formed; it is shown in Table 8.

The initial diagonal estimates for Q_1 might as well be the same first approximations that have served so well with centroid factoring—the highest non-diagonal entry in each of the columns. These need not be taken from

within Q_1. They can come from the first three rows of R_1, as they do for columns 4, 5, and 6. The resulting variable part $(G'E_1G)$ of (15) appears in Table 9. The subscript on E_1 indicates that it is the first in a series of E matrices.

The sum of the matrices in Tables 8 and 9 is given by Table 10. The non-vanishing side entries in Table 10 call attention to the fact that there is

TABLE 8
$G'Q_{10}G$

	A	B	C
A	.92	1.13	.83
B	1.13	.37	.24
C	.83	.24	1.61

TABLE 9
$G'E_1G$

	A	B	C
A	2.72	-1.17	-.63
B	-1.17	2.08	-.07
C	-.63	-.07	1.16

TABLE 10
$K_1^{-2}=G'Q_{10}G+G'E_1G$

	A	B	C
A	3.64	-.04	.20
B	-.04	2.45	.17
C	.20	.17	2.77

nothing in the equations which prevents this estimated K^{-2} from being non-diagonal when the data are imperfect. Some reduction of the side entries in K^{-2} may occur as the iterations proceed, but they may never vanish identically.

A first estimate of K^{-1} is obtained by taking the square roots of the diagonal entries in Table 10. Thus the side entries are not used. This suggests that a further shortening of the iterative procedure would be to compute only the diagonal terms in (15).

The first estimate of V_{a1}, obtained by (14), is shown in Table 11. The last column of Table 11 gives the new diagonal estimates (the sums of squared row entries in this first V_{a1}) which make up the E_2 of the second trial. Now the matrices $G'E_2G$, K_2^{-2}, K_2^{-1}, and CK_2^{-1} can be formed in turn exactly as in the first trial, the resulting second estimate of V_{a1} providing an E_3, etc. In the present example this process was continued for a total of

TABLE 11
First $V_{a1}=CK_1^{-1}$

	A	B	C	h_1^2
4	.50	-.02	-.08	.26
5	.13	.64	.23	.48
6	.08	-.09	.65	.44
7	.42	.42	.12	.37
8	.29	-.08	.53	.37
9	.17	.17	.53	.34
10	.42	.28	.28	.33

TABLE 12
Final $V_{a1}=CK_4^{-1}$

	A	B	C	h_4^2
4	.50	-.02	-.08	.26
5	.14	.66	.24	.51
6	.08	-.10	.66	.45
7	.42	.44	.12	.38
8	.29	-.08	.54	.38
9	.17	.18	.54	.35
10	.42	.29	.29	.35

four trials, at the end of which E_5 became identical with E_4 at the level of accuracy being used, so that no further change in the estimated V_{a1} could take place. This final V_{a1} is shown in Table 12, along with the final diagonal estimates for Q_1. It will be seen that no essential change has occurred from the first to the final V_{a1}, although all diagonal estimates but the first have

changed. Most of the changes occurred between the first and second estimates of V_{a1}. All changes after the second V_{a1} (six loadings and two diagonals) were of size .01. This suggests that one repetition of the iterative procedure may often suffice for practical purposes.

Now V_{b1} is formed by means of (16) and is recorded in Table 13. Tables 12 and 13 together make up V_1, so that V_2 can now be obtained by (3). V_2 appears in Table 14. The structure of V_1 and V_2 is further removed than was anticipated from the simple structure configuration which generated this

TABLE 13
$V_{b1}=P_1'GK_4$

	A	B	C
1	.50	-.07	-.03
2	.17	.59	.20
3	.10	-.02	.59

TABLE 14
$V_2=V_1D$

	A	B	C
1	.67	-.07	-.02
2	.23	.63	.15
3	.13	-.02	.43
4	.67	-.02	-.06
5	.19	.70	.18
6	.11	-.11	.49
7	.56	.47	.09
8	.39	-.09	.40
9	.23	.19	.40
10	.56	.31	.21

example. Apparently the result is highly sensitive to such minor distortions as rounding error and the random increments that were mentioned earlier. Such small changes seem able to force a sizeable shift in the position of the reference frame before the maximum degree of column proportionality is restored. This apparent instability in rotation should perhaps be kept in mind in the application of any proportional profiles or latent structure solution.

Further rotation of V_1 and V_2 would improve their simple structure appearance, but that would of course reduce the extent of their column proportionality. This illustrates the fact that nothing in the equations for proportional profiles guarantees a simple structure solution. The proportional profiles result might be nearer to a centroid or principal components analysis. However, it would seem that if the notion of proportional profiles is to be generally applicable and psychologically meaningful, then the required differential selection should take place with respect to something like the factors of simple structure, where factorial composition is at least partially independent of the make-up of the particular test battery.

The degree of fit for the present example is indicated by the two residual matrices shown in Tables 15 and 16. The difference in goodness of fit for R_1 and for R_2 calls attention to a consideration which is purely empirical. With perfect data it would make no difference which of the two correlation matrices was designed R_1 and which was called R_2. With empirical data, on the other

hand, such matters as sampling and sample size may appropriately influence the choice. It is to be expected that R_1 will be fitted best, since P_1 and Q_1 are involved in the solutions for C and K^{-1}. It is also to be expected that the P_2 section of R_2 will be fitted better than the rest of R_2, since P_2 is the only part of R_2 that is used in the solution. If it were considered to be worth the trouble, the fit of the two correlation matrices could be more nearly equalized by post-multiplying C by D to form $V_{a2}K$. Then the iterative procedure

TABLE 15
$R_1 - V_1V_1'$

	1	2	3	4	5	6	7	8	9	10
1		.01	-.01	.00	.00	-.01	.00	.00	.00	.00
2	.01		-.02	.01	.00	-.01	.00	.00	.00	.01
3	-.01	-.02		.00	.00	-.01	.00	.00	.00	-.01
4	.00	.01	.00		.01	.01	.00	.01	.01	.00
5	.00	.00	.00	.01		.02	.01	.02	.01	.02
6	-.01	-.01	-.01	.01	.02		.01	-.01	.01	.00
7	.00	.00	.00	.00	.01	.01		.02	.01	.01
8	.00	.00	.00	.01	.02	-.01	.02		.02	.01
9	.00	.00	.00	.01	.01	.01	.01	.02		.02
10	.00	.01	-.01	.00	.02	.00	.01	.01	.02	

TABLE 16
$R_2 - V_2V_2'$

	1	2	3	4	5	6	7	8	9	10
1		.02	.00	.01	.01	-.01	-.02	-.03	.00	.02
2	.02		-.03	.03	-.01	-.02	.01	.01	.02	.01
3	.00	-.03		.02	.01	-.03	-.03	.02	.00	-.01
4	.01	.03	.02		.00	.00	.01	.02	.01	.06
5	.01	-.01	.01	.00		.02	-.05	.04	.02	-.02
6	-.01	-.02	-.03	.00	.02		-.03	-.05	-.04	-.03
7	-.02	.01	-.03	.01	-.05	-.03		-.02	-.01	-.01
8	-.03	.01	.02	.02	.04	-.05	-.02		.00	.01
9	.00	.02	.00	.01	.02	-.04	-.01	.00		.01
10	.02	.01	-.01	.06	-.02	-.03	-.01	.01	.01	

could be applied to obtain a second estimate of K^{-1} based on the Q_2 section of R_2. Some sort of average (possibly the geometric mean) of the two K^{-1} estimates should then lead to a solution fitting the two correlation matrices to about the same extent. The Cattell solution for proportional profiles probably provides better over-all fit, but then R_1 must be factored and the communality problem reappears.

It may be mentioned, in passing, that the usual summational checks are applicable at nearly every stage of the present solution. They have been omitted here to simplify the exposition.

REFERENCES

1. Anderson, T. W. On estimation of parameters in latent structure analysis, *Psychometrika*, 1954, 19, 1–10.
2. Birkhoff, G., and MacLane, S. A survey of modern algebra. New York: Macmillan, 1953.
3. Cattell, R. B. "Parallel proportional profiles" and other principles for determining the choice of factors by rotation. *Psychometrika*, 1944, 9, 267–283.
4. Cattell, R. B., and Cattell, A. K. S. Factor rotation for proportional profiles: analytical solution and an example. *Brit. J. statist. Psychol.*, 1955, 8, 83–92.
5. Gibson, W. A. An extension of Anderson's solution for the latent structure equations. *Psychometrika*, 1955, 20, 69–73.
6. Green, B. F. A general solution for the latent class model of latent structure analysis. *Psychometrika*, 1951, 16, 151–166.
7. Lazarsfeld, P. F. The logical and mathematical foundation of latent structure analysis. Chapter 10 in Stouffer, S. A. et al.: Measurement and prediction. Princeton: Princeton Univ. Press, 1950.
8. Thurstone, L. L. Multiple-factor analysis. Chicago: Univ. Chicago Press, 1947.

Manuscript received 5/4/54

Revised manuscript received 3/10/55

PSYCHOMETRIKA—VOL. 21, NO. 2
JUNE, 1956

A NON-PARAMETRIC TEST OF CORRELATION USING RANK ORDERS WITHIN SUBGROUPS

WARREN S. TORGERSON

EDUCATIONAL TESTING SERVICE

Kendall's rank order test for association between two variables is generalized to the case where the total sample is made up of several subgroups and the data on one or both variables consist of the rank order within each subgroup. The test involves no assumptions concerning scales of measurement, shapes of distributions, or relative level of excellence or amount of variability of the different subgroups. Two empirical examples indicate that the normal approximation to the exact test of significance can be considered adequate for most practical situations. Special consideration is given to the case of tied ranks. If ties occur in but one variable within any given subgroup, only a slight modification in procedure is needed. Extensive ties in both variables within subgroups lead to difficulties in determining the appropriate correction for continuity.

I. *Problem*

The problem occasionally arises of determining whether or not a significant correlation exists between two variables when the total sample is made up of a number of subsamples, and the scores on one or both variables consist of rank orders within the subsamples. Such a situation frequently occurs, for example, when one wishes to determine whether or not scores on a psychological test are significantly related to supervisors' rankings with respect to proficiency on some task. Similar situations commonly occur in the military establishment. In the measurement of personality traits it is also desired at times to determine the significance of correlation between a trait measured by a personality schedule or test and ratings or rankings on the trait as made by peers. In any of these situations it frequently happens that the population is best considered as composed of many subgroups, each headed by its own supervisor or officer, or composed of subgroups whose members are well acquainted with one another.

Let us consider the first situation in somewhat more detail. Assume that a total sample is composed of k small subgroups (of from, say, three to six or seven subjects each), each headed by a supervisor. Each supervisor can place his own subordinates in rank order with respect to proficiency on the task of interest. No supervisor, however, has any information concerning proficiencies of subjects supervised by others. Hence, no information is available concerning the relative positions of subjects from different subgroups. It is not known, for example, whether the highest ranked person in one

subgroup is better or worse on the criterion than the lowest ranked person in another subgroup.

We thus have as raw data, first, scores for all subjects on the psychological test, and, second, criterion data in the form of k sets of ranks, one set for each of the k subgroups composing the total sample. The problem is to determine whether or not a significant correlation obtains between the test and the criterion.

When one is confronted with this problem, perhaps his usual procedure is either to decide not to carry out the experiment after all or to use procedures which require one or another of several sets of simplifying assumptions.

One such procedure, for example, would be to throw out enough observations so that each subgroup contains the same number of cases, treat the ranks as if they were a single rank order with k ties at each rank, and then test the significance of the rank correlation. Rather than eliminating cases to equate the size of subgroups, linear transformations could be made on the original ranks in such a way as to equate the average rank and the spread of the ranks for the different subgroups, and the separate ranks then combined into a single rank order. Either of these procedures involves a rather imposing set of assumptions.

A different approach would be to compute a test of significance for each subgroup separately and then combine the tests. Either a binomial test or a chi-square test using the $-2 \log p$ transformation might be used (1). However, due to the extremely small numbers of cases within the separate subgroups, neither would seem especially appropriate here: the binomial test because it uses only a portion of the data, and the chi-square test because the highly discrete nature of the data violates the assumption of a rectangular continuous distribution of probabilities.

The purpose of this note is to present a simple over-all non-parametric test for correlation between the two variables. No assumptions whatsoever are made concerning scales of measurement, shapes of distributions, or relative level of excellence and/or amount of variability between subgroups.

II. *The Test of Significance*

The proposed test is an extension of Kendall's test of association between two rank orders (2). His test involves, first, computation of the Kendall sum and, second, the determination of the probability of obtaining a sum that large or larger by chance alone. If the individuals are placed in rank order with respect to one of the variables, the Kendall sum is the number of pairs of individuals in the same order on the second variable minus the number of pairs of individuals in the reverse order. (It should be noted that special rules for computation make actual counting of all pairs unnecessary.) Since there are $n(n-1)/2$ possible pairs for n subjects, the possible values of the sum range from $-n(n-1)/2$ through $+n(n-1)/2$. The distribution of

possible values of the sum under the null hypothesis of zero correlation between the two variables is symmetrical about zero and has a variance of $n(n-1)(2n+5)/18$. Further, the distribution approaches normality very rapidly, so that with 10 or more subjects, the normal approximation gives adequate results. Exact distributions have been computed for n's up to 10 (2, p. 141).

The Kendall sum, of course, deals with rank orders of two variables for a single group. What then of our problem concerning several subgroups? Here we find that the Kendall sum has some exceedingly pleasing properties.

Let s_j denote the Kendall sum of the jth subgroup, where $j = 1, 2, \cdots, k$. We can obtain a *sum* of Kendall sums for the k subgroups. Let S denote this sum:

$$S = \sum_{j=1}^{k} s_j \,. \tag{1}$$

The distribution of S under the null hypothesis of zero correlation within each subgroup is also symmetrical about zero and has a variance equal to the sum of the variances of the sums (σ_j^2) of the separate subgroups,

$$\sigma_S^2 = \sum_{j=1}^{k} \sigma_j^2 = \sum_{j=1}^{k} n_j(n_j - 1)(2n_j + 5)/18. \tag{2}$$

Further, the distribution also appears to approach normality very rapidly. Using the normal approximation, the test is simply a test of the significance of the critical ratio

$$\mathrm{CR} = \frac{S \mp 1}{\sigma_S} = \frac{\sum_{j=1}^{k} s_j \mp 1}{\sqrt{\frac{1}{18} \sum_{j=1}^{k} n_j(n_j - 1)(2n_j + 5)}}. \tag{3}$$

The "one," which is subtracted from S whenever S is positive and added to S whenever S is negative, is a correction for continuity. Since possible values of S in any given situation are either all odd numbers or all even numbers with a class interval of two, the correction for continuity, taken as one half the class interval, is unity.

To get an idea of whether or not the distribution approaches normality rapidly enough to be of any use, two empirical tests were run—one using subgroups of sizes 3, 4, 4, and 6, and the other using subgroups of sizes 4, 4, 5, and 5. For every possible value of S the exact probability of obtaining a value of S that large or larger was computed, along with the corresponding probability using the normal approximation. The results are indicated in Table 1. As can be seen, the approximation is excellent even with samples as small as these. Since most practical situations will involve either more subsamples or subsamples of greater size, the normal approximation can be considered adequate.

TABLE 1

Comparison of Probabilities Computed Using the Normal Approximation (P_N) with the Corresponding Exact Values (P_E)*

Example 1			Example 2		
n_j = 3, 4, 4, and 6			n_j = 4, 4, 5, and 5		
S	P_N	P_E	S	P_N	P_E
			32	.0000	.0000
30	.0000	.0000	30	.0000	.0000
28	.0001	.0000	28	.0001	.0000
26	.0002	.0000	26	.0002	.0001
24	.0005	.0002	24	.0006	.0003
22	.0014	.0008	22	.0016	.0010
20	.0034	.0024	20	.0038	.0029
18	.0078	.0064	18	.0085	.0073
16	.0164	.0149	16	.0175	.0164
14	.0321	.0311	14	.0339	.0331
12	.0587	.0588	12	.0612	.0612
10	.1001	.1016	10	.1031	.1042
8	.1594	.1623	8	.1628	.1649
6	.2383	.2417	6	.2414	.2438
4	.3347	.3374	4	.3368	.3388
2	.4435	.4445	2	.4443	.4449

*Only the positive half of the distribution is tabulated, the negative half being obtained by symmetry. P refers to the probability of obtaining a value of S that large or larger under the null hypothesis.

III. *Numerical Example*

We shall apply the test to the set of hypothetical data given in Table 2. The total sample is made up of five subgroups ($j = 1, 2, \cdots, 5$) of 3, 5, 6, 8, and 8 subjects respectively. For each subject in each subgroup, two scores are available, the score on variable 1 being expressed in terms of the rank position of the subject within his own subgroup and the score on variable 2

TABLE 2

Raw Data for Numerical Example*

Group 1		Group 2		Group 3		Group 4		Group 5	
V_1	V_2	V_1	V_2	V_1	V_2	V_1	V_2	V_1	V_2
1	75	1	46	1	60	1	67	1	63
2	62	2	54	2	77	2	38	2	70
3	56	3	65	3	51	3	41	3	68
		4	40	4	47	4	44	4	59
		5	32	5	56	5	43	5	48
				6	49	6	37	6	67
						7	59	7	51
						8	30	8	54

*Scores on variable 1 (V_1) are expressed in terms of rank order within subgroup and on variable 2 (V_2) in terms of test scores.

expressed in terms of ordinary test scores. The subjects have been arranged in order of their rank position within their own subgroup.

The first step is to transform the test scores into ranks within subgroups. This has been done in Table 3. We now compute the Kendall sum for each subgroup. Since the subjects in each subgroup have been ordered with respect to variable 1, this is accomplished most easily by first determining for each subject in the subgroup the number of subjects below him who have a higher rank number on variable 2. For example, in subgroup 2 there are two subjects with higher rank numbers below the first subject, two below the second subject, two below the third subject, one below the fourth, and none below the fifth. If the sum of these numbers is denoted by r, then

$$s = 2r - \tfrac{1}{2}n(n - 1).$$

For subgroup 2,

$$r = 2 + 2 + 2 + 1 = 7,$$

and, hence,

$$s_2 = 14 - \tfrac{1}{2}(5)(4) = 4.$$

In like manner, $s_1 = 3$, $s_3 = 7$, $s_4 = 8$, and $s_5 = 12$, and the sum of Kendall sums, $S = 3 + 4 + 7 + 8 + 12 = 34$.

TABLE 3

Computation of Values of s_j and $\underline{S}$*

Group 1		Group 2		Group 3		Group 4		Group 5	
V_1	V_2	V_1	V_2	V_1	V_2	V_1	V_2	V_1	V_2
1	1	1	3	1	2	1	1	1	4
2	2	2	2	2	1	2	6	2	1
3	3	3	1	3	4	3	5	3	2
		4	4	4	6	4	3	4	5
		5	5	5	3	5	4	5	8
				6	5	6	7	6	3
						7	2	7	7
						8	8	8	6

$r_1 = 2+1 = 3$

$r_2 = 2+2+2+1 = 7$

$r_3 = 4+4+2+0+1 = 11$

$r_4 = 7+2+2+3+2+1+1 = 18$

$r_5 = 4+6+5+3+0+2+0 = 20$

$s_1 = 6 - 3 = 3$

$s_2 = 14 - 10 = 4$

$s_3 = 22 - 15 = 7$

$s_4 = 36 - 28 = 8$

$s_5 = 40 - 28 = 12$

$\underline{S} = 34$

*Scores on both variables expressed in terms of rank order within subgroup.

The variance of S under the null hypothesis is equal to the sum of the variances of the five separate subgroups. From equation 2 we have

$$\sigma_S^2 = \tfrac{1}{18}(66 + 300 + 510 + 1176 + 1176) = 179.3333.$$

From (3),

$$\mathrm{CR} = \frac{34 - 1}{\sqrt{179.3333}} = 2.464$$

which, using a two-tailed test, is significant beyond the .02 level.

IV. *Tied Ranks*

Heretofore we have assumed implicitly that ties would not occur. When ties do occur several alternative procedures are available. One such procedure would be to assign ranks within ties randomly and then proceed as if no ties had occurred. A second procedure might be to assign ranks within ties in the most favorable way and also in the least favorable way and then run a separate test of significance on each. If both tests lead to the same conclusion all is well. However, this procedure is inadequate if the two tests lead to conflicting decisions.

A third procedure is to assign the mean rank to the ties. This is the procedure adopted by Kendall for the single group case.

The mean rank method, or mid-rank method as it has been called, requires some additional considerations in computation of both the Kendall sum and its variance for the individual subgroup. Two cases need to be distinguished: (*i*) that in which ties occur, for any given subgroup, in only one of the two variables; (*ii*) that in which ties occur in both variables within a subgroup.

(*i*) *Ties in one variable only for any given subgroup*

It should be noted at the onset that it is irrelevant whether the variable containing the ties is the same variable in all subgroups, or whether for some subgroups it is one variable and for others, the other variable.

The Kendall sum, s_j for any subgroup j has the same meaning as before with the additional stipulation that tied pairs, being neither in the same nor in the reverse order on the second variable, are counted as zero. If the subjects are arranged in order on the variable containing no ties, the sum is the number of pairs on the second variable that are in the same order minus the number of pairs in the reverse order. Tied pairs neither add to nor subtract from the sum.

Calculation of the variance of the jth subgroup requires the following modification. There may be several different ties, and any number of subjects may be involved in any given tie. Let a be an index referring to a particular tie ($a = 1, 2, \cdots, g$) and t_{aj} refer to the number of subjects involved in the ath tie of the jth subgroup. Then Kendall has shown that

$$\sigma_j^2 = \frac{1}{18}\left[n_j(n_j - 1)(2n_j + 5) - \sum_{a=1}^{g} t_{aj}(t_{aj} - 1)(2t_{aj} + 5)\right]. \qquad (4)$$

The sum of Kendall sums and its variance under the null hypothesis remain the sum of the individual sums and the sum of the individual variances, respectively. The distribution remains symmetrical about zero and also appears to approach normality rapidly. Possible values of S for any particular combination of number and size of subgroups remain either all odd or all even, with a step interval of two. Hence for the normal approximation, the

correction for continuity remains one, and the critical ratio,

$$\mathrm{CR} = \frac{\sum_{j=1}^{k} s_j \mp 1}{\sqrt{\sum_{j=1}^{k} \sigma_j^2}}, \tag{5}$$

where the individual subgroup sums and variances are computed as noted above.

(ii) Ties in both variables for some or all subgroups

The computation of the separate Kendall sums is the same as before except that ties in *either* rank do not contribute to the sum (see **2**, p. 26, for more detail). The formula for the separate variances is given below, where t_{aj} refers to one variable and u_{bj} to the other.

$$\begin{aligned}\sigma_j^2 = \tfrac{1}{18}[&n_j(n_j - 1)(2n_j + 5) \\ &- \sum_a t_{aj}(t_{aj} - 1)(2t_{aj} + 5) - \sum_b u_{bj}(u_{bj} - 1)(2u_{bj} + 5)] \\ &+ \frac{1}{9n_j(n_j - 1)(n_j - 2)}[\sum_a t_{aj}(t_{aj} - 1)(t_{aj} - 2)] \\ &\qquad \cdot [\sum_b u_{bj}(u_{bj} - 1)(u_{bj} - 2)] \\ &+ \frac{1}{2n_j(n_j - 1)}[\sum_a t_{aj}(t_{aj} - 1)][\sum_b u_{bj}(u_{bj} - 1)].\end{aligned} \tag{6}$$

Again S is the sum of the separate sums, and its variance the sum of the separate variances. The distribution is again symmetrical about zero. However, the correction for continuity depends upon the particular situation. While in the extreme case (a single subgroup with both variables dichotomized) the appropriate correction gets as high as $n/2$, it may be surmised that, for practical situations where there are numerous subgroups and but very few ties occurring in both ranks within the same subgroup, the unit correction for continuity will not be too bad an approximation. An alternative procedure would be to convert the problem to the "ties in one variable only for any subgroup" case by assigning a rank order randomly to the ties in one variable.

REFERENCES

1. Jones, L. V., and Fiske, D. W. Models for testing the significance of combined results. *Psychol. Bull.*, 1953, 50, 375–382.
2. Kendall, M. G. Rank correlation methods. London: Griffin, 1948.

Manuscript received 6/6/55

PSYCHOMETRIKA—VOL. 21, NO. 2
JUNE, 1956

MAXIMIZING TEST VALIDITY BY ITEM SELECTION

HAROLD WEBSTER

MELLON FOUNDATION

VASSAR COLLEGE

The exact condition for discarding k items from a test in order to obtain a residual test with higher validity is derived. A proof that validity always increases is given for the case $k = 1$. The lack of uniqueness of maximum validity when achieved by use of the condition is discussed. With the use of additional restrictions on items to be included in the initial test, a practical test construction procedure which has several advantages over previous methods is developed. The homogeneity of tests constructed by the method is discussed, and applications are given.

The problem of selecting from among a large number of test items those yielding a test score which will correlate maximally with an external variable has been studied by a number of writers. Gleser and DuBois (6) and Gulliksen (8) have summarized pertinent research. The problem is of first importance when it is necessary to construct a test which has a high correlation with some known criterion; the same problem would arise, however, if it were necessary to select those parts of a composite experimental criterion which would correlate highest with a fixed test. Mathematical solutions to either problem, or to the simultaneous solution of both problems (10), are known. These solutions are impractical in the case of item selection, however, because of the large number of partial regression weights required for items.

An interesting method of testing composites for sampling stability is known (3) but its application for comparing numerous item composites would present serious difficulties. Recently Lord (12) has derived sampling variances for test statistics under conditions of item sampling, but presumably the behavior of the validity coefficient when there is sampling of both persons and items is still unknown. Guilford (7) advises that the approximate solutions be applied only with large tests which have been administered to large numbers of subjects. Richardson and Adkins (13) suggest that possibly any item selection index would be so susceptible to sampling fluctuations that a choice among methods of selecting items would be practically a matter of no importance. Tukey (15) includes item selection for validity among the important unsolved problems of experimental statistics.

Because of difficulties inherent in exact solutions, and despite lack of precise knowledge of sampling fluctuations of item composites, investigators

have devised various approximation methods. Toops (14) advocated the construction of tests starting with one highest correlating item and adding, one at a time, items which most increased the validity of the composite. Adkins has applied this method and she and Richardson (13) found that a much less laborious modification also gave good results. They selected from a test a number of items at once with highest criterion partial regression weights, the test comprising the other independent variable. The approximate item weight used requires more computing than the one derived in the present paper. Most of the earlier methods, including Toops', assumed a unique order for the item selection; for example, selected items were not reexamined at a later stage to see if they still belonged in the regression structure. Because of random fluctuation in subsequent samples what is gained by such reexamination may indeed not be worth the additional labor; but if the method of reexamination were brief enough, it would seem worth while to use it.

Horst (9) advises starting with a relatively large test and rejecting those items for which the ratio of a function of item-criterion covariance to a function of item-test covariance is small. The items are plotted and then selected using the functions as coordinates. Horst cautions against discarding too many items at once without recalculating item-test parameters. Gulliksen (8) has simplified Horst's procedure somewhat, but it would still seem better to avoid plotting if a more economical method of identifying items to be rejected can be found.

Flanagan (4) constructed tests by retaining those items for which the item-criterion correlation exceeded the item-test correlation; he advised repeating the process, but did not indicate whether he thought it advisable to reexamine previously rejected items to see whether they should be put back into the test. Gleser and DuBois (6) recommend what is essentially Flanagan's procedure for a first approximation, and a more refined one for subsequent iterations. It is unclear how they would treat items with positive validities and negative test correlations; rejection of such items always decreases test validity. Their suggestion that the initial test be restricted to positively valid items seems to be a good one, but selection conditions were not derived under this restriction. Gleser and DuBois also state some conditions under which item variances may be ignored in the first approximation to the final test; with the method of the present paper item variances are not required.

In the present paper the exact item selection condition is derived, and the approximations which are used can be seen to be close enough and at the same time to require less computation in applications than any previous ones. Also, a study of the selection condition itself reveals why the traditional expedient of imposing a validity condition on *individual* items to be admitted into a test is reasonable.

I. *Derivation and Properties of the Exact Condition for Discarding k Items in Order to Obtain a Residual Test of Higher Validity*

In order to construct a test with high validity, first a large number n of potential test items is examined for correlation with the external criterion C. It is then possible, with no loss of generality, to score each item j so that its covariance C_{jc} with the criterion C is positive or zero. The sum of these n item scores, $x_1 + \cdots + x_i + \cdots + x_n$, forms an experimental test T from which it is desired to discard items such that the residual test M correlates maximally with C. The correlation of T with C may be written

$$r_{TC} = \sum_i^n C_{ic}/S_T S_C = C_{TC}/S_T S_C , \tag{1}$$

where C's are covariances, S's standard deviations, and the summation is over the n item scores $1, \cdots, i, \cdots, n$. If the first k items, designated $1, \cdots, j, \cdots, k$, are discarded from T, the validity of the residual test M is

$$r_{MC} = (C_{TC} - \sum_i^k C_{ic})/S_M S_C . \tag{2}$$

Since M is to be more valid than T, the condition

$$r_{TC} < r_{MC} \tag{3}$$

must be satisfied. Using (1) and (2), (3) may be written

$$C_{TC}/S_T S_C < (C_{TC} - \sum^k C_{ic})/S_M S_C . \tag{4}$$

Making the substitution $S_{T-k} = S_M$, and multiplying by S_C, (4) becomes

$$C_{TC}/S_T < (C_{TC} - \sum^k C_{ic})/S_{T-k} . \tag{5}$$

As a consequence of having scored items so that $C_{ic} \geq 0$, the quantity in parentheses in (5) must, like the other terms, be positive. The terms of (5) can be rearranged as follows,

$$\begin{aligned} S_{T-k}/S_T &< 1 - [(\sum^k C_{ic})/C_{TC}], \\ (\sum^k C_{ic})/C_{TC} &< 1 - [S_{T-k}/S_T], \\ S_T/C_{TC} &< (S_T - S_{T-k})/\sum^k C_{ic} . \end{aligned} \tag{6}$$

Writing (6) in a form which will be more useful later,

$$C_{TC}/S_T > (\sum^k C_{ic})/(S_T - S_{T-k}). \tag{7}$$

The practical difficulty encountered in using (7) as a condition to be satisfied by rejecting k items in order to enhance validity lies in the fact that

the expansion of S_{T-k} contains the sum of inter-item covariances of the k items. Use of inter-item relationships is laborious and seldom feasible in the typical test construction situation where there is available at most a machine for counting item responses. Further study of (7) is also warranted because it is not clear whether any method exists for identifying uniquely a set of k items to be rejected.

Squaring (5), substituting $V_{T-k} = V_T + V_k - 2C_{kT}$, and simplifying gives

$$C_{TC}^2(V_k - 2C_{kT}) < V_T \sum^k C_{iC}(\sum^k C_{iC} - 2C_{TC}). \tag{8}$$

The right side of (8) will always be negative (all $C_{iC} \geq 0$), and it can be seen that items for which the left side of (8) is positive will not satisfy (3) or the subsequent conditions. For since the terms other than the numerator on the right in (6) are positive, the k items (including the case $k = 1$) may not be discarded by (6) with a resulting increase in validity unless $S_T > S_{T-k}$. But $V_k - 2C_{kT}$ in (8) is equal to $V_{T-k} - V_T$, a negative quantity when $S_T > S_{T-k}$. Multiplying (8) by -1 and rearranging terms, we obtain

$$\frac{C_{TC}^2}{V_T} > \frac{\sum^k C_{iC}(2C_{TC} - \sum^k C_{iC})}{2C_{kT} - V_k}, \qquad 2C_{kT} > V_k. \tag{9}$$

Although (9) is the general condition (under the restriction, $C_{iC} \geq 0$, from which it follows that $2C_{kT} > V_k$) for rejecting k items from T in order to obtain a more valid test, the case where $k = 1$ is more useful. When k is a single item j, (9) reduces to

$$\frac{C_{TC}^2}{V_T} > \frac{C_{jC}(2C_{TC} - C_{jC})}{2C_{jT} - V_j}, \qquad 2C_{jT} > V_j. \tag{10}$$

The second Gleser and DuBois condition referred to previously is a fairly good approximation of (10). It can be obtained from (7) when k is a single item j [another way of writing (10)] by dividing (7) by S_T and substituting $S_T S_{T-j} \doteq (V_T + V_{T-j})/2$; (7), (9), and (10) have recurrence properties not possessed by any approximations known to the writer.

A proof follows that each successive application of (7) when $k = 1$, that is, when single items are discarded successively, increases validity. It can be generalized immediately for $k > 1$. The first inequalities in the recursion series for (7), when single items 1, 2, $\cdots$ are rejected, are

$$\frac{C_{TC}}{S_T} > \frac{C_{1C}}{S_T - S_{T-1}}, \qquad \frac{C_{(T-1)C}}{S_{T-1}} > \frac{C_{2C}}{S_{T-1} - S_{T-2}}, \qquad \cdots. \tag{11}$$

If the terms on the left of the inequality signs in sequence (11) were divided by the constant S_C, they would become the correlations which are the validity coefficients as the test becomes shorter. These terms therefore comprise a

sequence with a maximum upper bound equal to S_C. Every bounded monotone sequence is convergent, and

$$C_{TC}/S_T\,, \qquad C_{(T-1)C}/S_{T-1}\,, \qquad \cdots \tag{12}$$

is bounded and can be shown to be monotone. Given that the first inequality of (11) holds for item 1, we first prove that

$$C_{TC}/S_T < C_{(T-1)C}/S_{T-1}\,. \tag{13}$$

By substituting $C_{(T-1)C} = C_{TC} - C_{1C}$ and clearing of fractions, (13) can also be written

$$S_{T-1}C_{TC} < S_TC_{TC} - S_TC_{1C}\,. \tag{14}$$

But this is another way of writing the first term of (11), and this proves that discarding item 1 has increased validity from r_{TC} to $r_{(T-1)C}$. By induction, if subsequent items are rejected by (11), the validity always increases. The sequence (12) is therefore monotone and convergent. By comparison with (12) the sequence

$$\frac{C_{1C}}{S_T - S_{T-1}}\,, \qquad \frac{C_{2C}}{S_{T-1} - S_{T-2}}\,, \qquad \cdots\,, \tag{15}$$

which is formed by the terms on the right of the inequalities in sequence (11), is also convergent but not necessarily monotone. Conditions can be written which will make (15) monotone, but the writer was unable to find any which were practical and at the same time would insure unique maximum validity.

Nevertheless, (11) converges as it stands. (A difference between two convergent sequences also converges). This is sufficient reason to make use of an analogous sequence, one comprised of terms which are approximations of (10), as an aid in constructing valid tests.

II. *Development of a Practical Test Construction Procedure*

Since rejection of a unique set of items from a test in order to maximize (obtained) validity for a given sample is in general not possible, and since the sampling variance of the validity when both items and persons are sampled is unknown, the need for item selection restrictions in addition to (10) is obvious.

A practical restriction traditionally used in test construction is that items should each have enough variance to share appreciably in the discrimination of subjects. A second restriction, that items retained in the test have individual validity, follows from (6). Reference to (6) when k is a single item j shows that because items are scored so that $C_{jC} \geq 0$, discarding items for which $2C_{jT} - V_j \equiv V_T - V_{T-j} < 0$ will not increase test validity. Because of the scoring convention, therefore, the sign of $S_T - S_{T-j}$ is de-

pendent upon the sign of C_{iC}, which in turn is subject to sampling fluctuations. In order to be reasonably confident that these signs, and consequently the test scoring, will not vary in subsequent samples, it is necessary that most of the items in the original test be valid. This condition alone could be satisfied by requiring that each item-criterion correlation significantly exceed zero, or that

$$C_{iC} \geq S_i S_C z/\sqrt{N}, \tag{16}$$

where z is a chosen critical value for a normal deviate, say 1.96.

Setting $S_i = 0.5$, the maximum value, in (16) will provide a conservative test for all items for which $S_i < 0.5$. That is, items satisfying (16) when $S_i = 0.5$ are significantly correlated with C at the level specified by z, but if a low-variance item is to be included in the test, then this must be compensated for by a higher covariance with the criterion. Including in the initial test T only items satisfying (16) when $S_i = 0.5$, therefore, has the desired effect of insuring both that the initial test will contain valid items, and that most of them will have large variances. Rejection of items from T can then be started by applying an approximation of (10).

Ideally (10), or an approximation, would be recomputed after any item had been discarded by it before discarding another item. But rejecting items one at a time is not feasible because of the labor of recomputing values for the C_{iT}. A workable procedure, which is essentially the same as that used by Gleser and DuBois, is first to reject *all* items for which (10) holds (without recomputing C_{iT}) and then to reexamine the rejected items to see which would logically be put back into the test. The new test so formed may then be treated as if it were an initial test and the process continued until there are no further increases in validity.

In order to secure a suitable approximation of (10), first note that C_{iC}^2 is the final term in the expanded numerator on the right, and that C_{TC}, V_i, and C_{iC}^2 are in descending orders of magnitude. Substituting for the final C_{iC} its mean value in T, $C_{iC} \doteq C_{TC}/n$, and substituting the *maximum* value for V_i, 0.25, reduces (10) to

$$\left(\frac{2n}{2n-1}\right)\frac{C_{TC}}{V_T} > \frac{C_{iC}}{C_{iT} - 0.125}, \tag{17}$$

where n is the number of items in T. The use of the approximation $V_i \doteq 0.25$ is justified because, first, after applying (16) as described, most item variances are large; second, marginal items which would ordinarily be retained by (10) despite small variances will not be discarded by (17); and finally, the exactness of the value for V_i in (17) is increasingly unimportant with increasing n.

Suppose that m items for which (17) holds have been discarded, leaving a residual test M. Since (17) is progressively less accurate as items beyond the first are discarded, it is likely that some of the discarded items should be

put back into the test. Any rejected item j which satisfies the validity condition

$$r_{MC} \leq r_{(M+j)C} \tag{18}$$

will be added to test M. By a development similar to that used in deriving (10), (18) may be written

$$\frac{C_{MC}^2}{V_M} \leq \frac{C_{jC}(2C_{MC} + C_{jC})}{2C_{jM} + V_j}. \tag{19}$$

Using the approximations for C_{jC}^2 and V_j described above, (19) becomes

$$\left(\frac{2m}{2m+1}\right)\frac{C_{MC}}{V_M} \leq \frac{C_{jC}}{C_{jM} + 0.125}, \tag{20}$$

where m is the number of items in M.

Items for which (20) holds are added to M to form a third test. It is then possible to treat this third test as if it were T, the initial test, and again apply conditions (17) and (20); this process may be continued until no further increase in validity occurs.

It is also desirable to simplify computation of the covariances in (17) and (20). This can be done as follows: Test and criterion distributions are first transformed to comprise only five symmetrical categories containing the 8 per cent highest, 18 per cent next highest, 48 per cent middle, 18 per cent low, and 8 per cent lowest scores. A division of distributions into categories containing 9, 20, 42, 20, and 9 per cents of cases is recommended by Flanagan (5) as maximally efficient when scores in the categories are assigned values 2, 1, 0, − 1, and − 2, respectively. The distribution used in the present paper will be slightly less efficient than the one recommended by Flanagan, but will have the advantage of possessing a unit variance. Using only five categories, the covariances in conditions (17) and (20) may be obtained from item counts for the four extreme categories.

To transform (17) and (20) so that the covariances are of the form

$$C' = (2e + f - g - 2h)/N = D/N, \tag{21}$$

where e, f, g, and h are frequencies in the categories in the order given above (scores with zero weights, which are in the center category, are ignored), first note that the variance of every such forced distribution is a constant, $V' = 1.00$. Then if transformed values are indicated by primes, $S_C' = S_T' = 1$, and

$$r_{TC} = C_{TC}/(S_T S_C) = C_{TC}'/(S_T' S_C') = C_{TC}'. \tag{22}$$

Only one transformed distribution corresponding either to r_{jC} or to r_{jT} is needed; thus it is assumed that

$$r_{jT} = C_{jT}/(S_j S_T) = C_{jT}'/(S_j S_T') = C_{jT}'/S_j, \tag{23}$$

and a similar expression can be written for r_{ic}. Transforming (17) gives

$$\left(\frac{2n}{2n-1}\right)C'_{TC} > \frac{C'_{ic}}{C'_{iT} - (0.125/S_T)}, \tag{24}$$

and transforming (20),

$$\left(\frac{2m}{2m+1}\right)C'_{MC} \leq \frac{C'_{ic}}{C'_{iM} + (0.125/S_M)}. \tag{25}$$

A more convenient form of (24) is obtained by substituting $\sum^N T'C'/N = C'_{TC}$ and, from (21), $D/N = C'$ to get

$$\left(\frac{2n}{2n-1}\right)\frac{\sum^N T'C'}{N} > \frac{D_{ic}}{D_{iT} - (0.125N/S_T)}, \tag{26}$$

where n is the number of items in T. Similarly (25) is conveniently written

$$\left(\frac{2m}{2m+1}\right)\frac{\sum^N M'C'}{N} \leq \frac{D_{ic}}{D_{iM} + (0.125N/S_M)}, \tag{27}$$

where m is the number of items in M. Very little loss of accuracy occurs in (27), especially when $m/n > .8$ and $m > 30$, if the value of S_M is taken to be mS_T/n.

The procedure for applying (16), (26), and (27) is outlined in the next section.

III. *The Procedure for Applying the Item Selection Conditions*

1. Separate the answer sheets into five piles according to their criterion scores C, containing the 8 per cent highest, 18 per cent next highest, 48 per cent middle, 18 per cent low, 8 per cent lowest C scores, respectively. Mark papers in the four extreme C-score categories e, f, g, and h, respectively, so they can be identified in Step 6.

2. Record on item analysis sheets the values for e, f, g, and h, which are the frequencies of a response (for example, "true") in the highest, next highest, low, and lowest C-categories. Obtain from the item counts for the four extreme categories the difference

$$D_{ic} = 2e + f - (g + 2h) = NC'_{ic}$$

for each item j [see (21)]. Next choose the direction of response for each item so that $2e + f \geq g + 2h$, that is, so that $D_{ic} \geq 0$, and mark the items accordingly.

3. Apply (16) by including in the initial test T only those items for which $D_{ic} \geq 0.5z\sqrt{N}$, where z is the normal deviate corresponding to a chosen level of significance.

4. Score test T (with items scored in the directions determined in Step 2) and mark the scores on the answer sheets. Tally T and obtain its standard deviation S_T.

5. Separate the answer sheets into five piles as in Step 1, but this time according to their T scores.

6. Write in the frequencies for the cells of the 5×5 contingency table for the transformed T-scores and C-scores. Compute $\sum^N T'C'$, first noting that scores falling in the most extreme categories on either variable receive values $+2$ or -2, while those in next most extreme categories receive values $+1$ or -1. This can be done in a few minutes by counting, and frequencies for the center categories can be ignored.

7. Obtain the D_{jT} from item counts as in Step 2. Check for retention in the test any items for which $D_{jT} \leq 0.125N/S_T$. It is possible there may be no such items.

8. Compute the constant which is the left side of (26); set up the right side of (26) for each item j (the operations may not have to be carried out) and retain only those items for which (26) fails to hold.

9. Score the test M, comprised of items retained after Step 8. Tally M and obtain its standard deviation S_M, or use $S_M = mS_T/n$.

10. Separate the answer sheets into five piles as in Step 1, this time according to the M scores.

11. Obtain D_{jM} from item counts as in Step 3, *but only for those items previously discarded in Step 8.*

12. Obtain $\sum^N M'C'$ as in Step 6 and compute the left side of (27). Set up the right side of (27) for each item for which D_{jM} was obtained in Step 11, and mark items for which (27) holds to be put back into test M. This completes the first cycle of the iteration, and a large proportion of the possible increase in validity will have been obtained.

13. For convenience, again call the test obtained after Step 12 "test T," and repeat the procedure starting with Step 5. The iteration will stop at a point where either (26) or (27), applied alternately, will produce no further increase in test validity. Always apply (27) to all previously rejected items. If S' is a transformed score on the final test, $\sum^N S'C'/N$ [see (26) and (27)] is a conservative estimate of the final validity coefficient.

IV. *The Problem of Test Homogeneity*

A question which naturally arises is how much internal consistency tests will have when constructed by the method. Since the test becomes shorter and at the same time some of the redundant items with higher test correlations are dropped, its homogeneity may decrease. It will usually remain relatively high, however, especially if the initial test is long enough. Aside from length, another reason why homogeneity will be high (in practice between .82 and .90 for final tests of about 100 items) is that initial test

items are limited by (16) to those having significant criterion covariances; this condition alone tends to select items which correlate positively with the total test and thus enhance homogeneity.

Cronbach (2) has demonstrated the relative importance of high item-test correlations and test length in contributing to homogeneity. In symbolism of the present paper,

$$\frac{C_{iT}}{V_T} > \frac{V_i}{2}\left(\frac{1}{\sum^{n} V_i} - \frac{1}{V_T}\right), \qquad n > 7, \tag{28}$$

is a close approximation to $r_{(T+j)(T+j)} > r_{TT}$ where the latter are K.R. 20 coefficients for test T, including and excluding item j, respectively. Inequality (28) can be used to determine how large the item-test covariance should be before item j will contribute to the homogeneity of T. When (28) is transformed as in Part II, it is found, for tests and samples of only moderate size, that all items for which $D_{jT} > 1$ will contribute to homogeneity.

V. *Applications and Discussion*

The method is not time-consuming; once it is learned, two persons working with about 300 answer sheets, each of which contains as many as 540 true-false item responses, can construct a test in about 14 hours.

In the first two applications, the criterion variable was the score on an attitude test, the 20-item California Ethnocentrism Scale (1), and the tests constructed to correlate with it were selected from 379 true-false items from various standardized personality inventories. Items in the Ethnocentrism scale are hostile or disparaging statements about minority groups; each item receives a score from 1 to 7 to indicate extent of agreement.

In the first application using a sample of 288 college women, (16) was applied using $z = 1.96$ (see Part III, Step 3) with the result that 79 items were selected to comprise an initial test for which the validity, $\sum T'C'/N$, was .62. Of these, 14 were rejected by (26) leaving a 65-item test with a validity of .64. Applying (27) to the rejected items put one of them back, giving a 66-item test with a validity of .65. Subsequent applications of (26) and (27) resulted in rejections and selections of from 1 to 4 different items at a time accompanied by slight decreases in validity. The 66-item test was therefore accepted as the final test for this sample.

In the second application, using a sample of 50 middle-aged women, only 41 items from the 379 were selected by (16) for the initial test, even though z had been chosen because of the small sample size to correspond to the .10 level of significance. The initial validity for the 41 items was .77, a value undoubtedly largely spurious because of chance item-criterion correlations and because of the small sample size. Application of (26) discarded 9 items to increase validity to .80, and an application of (27) put 6 items back

into the test to make a 38-item test with validity of .82. This was not exceeded by subsequent iterations.

In several applications the necessity for using only items which were individually valid was demonstrated. In one case all positively scored items were used in the initial test, that is, condition (16) was not applied. The convergence was slow, and because of the presence of items which would be invalid in subsequent samples, the gain in test validity would not be expected to be permanent (see Part II). In another case only 58 items out of 677 available could be found which were related at the .05 level to grades of college freshmen. About 34 items would therefore be expected to have only a chance relationship to grades. Application of the method retained 41 items and raised validity from .52 to .60. But in a subsequent sample the validity was almost as small for the shortened test (.22) as for the initial test (.16). The shrinkage in both cases was obviously due to the large proportion of invalid items among the initial 58. These results show that it is necessary to insure item validity in the initial test before applying the rest of the method; this may be achieved either by applying (16) with z large when there is only one large sample, or by using several samples if z must be smaller.

If most of the items are valid, the method appears to be worth applying. For example, in a study reported elsewhere 178 items were found each to correlate at the .01 level with a criterion. The test validity was raised from .66 to .78 in the first sample ($N = 441$) by applying Flanagan's method (5), which is an approximation of the present method. A year later the shortened test (124 items) correlated .74 in a new sample ($N = 402$). In this case the difference .74 − .66 is significant, using the traditional z-transformation test, at the .03 level. Despite this apparently permanent gain in validity, the merit of selecting items for validity cannot be finally assessed until the appropriate sampling statistics are derived and applied.

REFERENCES

1. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., and Sanford, R. N. The authoritarian personality. New York: Harper, 1950.
2. Cronbach, L. J. Coefficient α and the internal structure of tests. *Psychometrika*, 1951, **16**, 297–334.
3. Fisher, R. A. The precision of discriminant functions. *Ann. Eugen., Lond.*, 1940, **10**, 422–429.
4. Flanagan, J. C. A short method of selecting the best combination of test items for a particular purpose. *Psychol. Bull.*, 1936, **33**, 603–4.
5. Flanagan, J. C. The effectiveness of short methods for calculating correlation coefficients. *Psychol. Bull.*, 1952, **49**, 342–348.
6. Gleser, G. C., and DuBois, P. H. A successive approximation method of maximizing test validity. *Psychometrika*, 1951, **16**, 129–139.
7. Guilford, J. P. Psychometric methods (2nd Ed.). New York: McGraw-Hill, 1954.
8. Gulliksen, H. Theory of mental tests. New York: Wiley, 1950. Chap. 21.
9. Horst, P. Item selection by means of a maximizing function. *Psychometrika*, 1936, **1**, 229–244.

10. Hotelling, H. Relations between two sets of variates. *Biometrika*. 1936, **28**, 321–377.
11. Loevinger, Jane, Gleser, Goldine C., and DuBois, P. H. Maximizing the discriminating power of a multiple-score test. *Psychometrika*, 1953, **18**, 309–317.
12. Lord, F. M. Sampling fluctuations resulting from the sampling of test items. *Psychometrika*, 1955, **20**, 1–22.
13. Richardson, M. W., and Adkins, Dorothy C. A rapid method of selecting test items. *J. educ. Psychol.*, 1938, **29**, 547–552.
14. Toops, H. A. The *L*-method. *Psychometrika*, 1941, **6**, 249–266.
15. Tukey, J. W. Unsolved problems of experimental statistics. *J. Amer. statist. Assoc.*, 1954, **49**, 706–731.
16. Webster, H. Approximating maximum test validity by a non-parametric method. *Psychometrika*, 1953, **18**, 207–211.

Manuscript received 3/3/55

Revised manuscript received 5/28/55

PSYCHOMETRIKA—VOL. 21, NO. 2
JUNE, 1956

THE VALIDITY OF THE SUCCESSIVE INTERVALS METHOD OF PSYCHOMETRIC SCALING*

WILLIAM W. ROZEBOOM AND LYLE V. JONES
UNIVERSITY OF CHICAGO

The degree to which scale values computed by the method of successive intervals diverge from theoretically "true" values is seen to be due to three types of error: error due to inequalities in variances of the distributions from which the scale values are computed, error due to non-normality of the distributions, and sampling error. The contribution of each type of error to the total error is evaluated; the latter is seen to be surprisingly small under appropriate conditions. Certain aspects of the formal methodology underlying scaling procedures are also briefly considered.

One of the most popular and perhaps the simplest of all methods by which stimuli can be assigned values for some psychological variable is the rating scale technique. Basically, a rating scale is some set of categories that partition sets of events into mutually exclusive classes. For example, a rating scale might be defined by the categories *high*, *medium*, and *low*, and a set of events generated by "the evaluation of the esthetic value of art object i by judge j," where i and j range over specified classes of art objects and judges. That is, each judge j assigns each art object i to a category of the rating scale, such an assignment constituting an event. Corresponding to each event designated by the coordinate pair i, j there is one and only one category: *high*, *medium*, or *low*.

Usually, the localization of each event on the scale is only a means to a representation of various subclasses of the events by a single value of the scale. This can be done by taking the most representative scale value of the distribution of scores in a subclass as the scale value for the subclass as a whole. Thus, for the example already given, we may be less concerned with the rating given to a particular art object by a given judge than we are in a rating representative of the values assigned to that object by the various judges. Since a specific art object defines a subclass of ratings, the most representative rating (however defined) can be taken as the value of this stimulus on the esthetic scale, not dependent upon any particular judge.

In its most elementary form, a rating scale imparts no measure of quantity to the events rated by it, merely being comprised of a set of mutually exclusive

*This paper reports research undertaken in cooperation with the Quartermaster Food and Container Institute for the Armed Forces, and has been assigned number 475 in the series of papers approved for publication. The views or conclusions contained in this report are those of the authors. They are not to be construed as necessarily reflecting the views or indorsement of the Department of Defense.

categories. An example of this simplest type is the standard color chart by which colors are classified. Stevens (9) has named this kind of scale a *nominal* scale. With appropriate additional data, assumptions, or definitions, however, the rating scale can be utilized as an ordinal, or even an interval scale.

If a relation can be obtained which orders the categories, then the rating scale has become an ordinal scale for that relation. One of the more customary ordering relations employed by psychologists in generating ordinal scales is that of preference, or choice. If the categories are such that, for a given judge, (*i*) an item assigned to category A is always chosen over any item assigned to category B (at least at the time of the assignment) and any item assigned to category B is always chosen over an item assigned to category C, and (*ii*) no item assigned to C is ever chosen over items assigned to A, then categories A, B, and C are ordered by the relation of preference.

It is customary at this point either to define or to hypothesize the existence of a psychological continuum underlying the categories of the rating scale, such that each category covers a range of the continuum, these ranges being exhaustive, mutually exclusive, and in the same ordinal relation as the corresponding categories. In short, the rating scale is interpreted as a gross technique by which the values of events are estimated on a similar, but much more discriminating underlying scale. Thus, art objects evaluated in terms of a three-category scale are assumed to be much more finely dis-

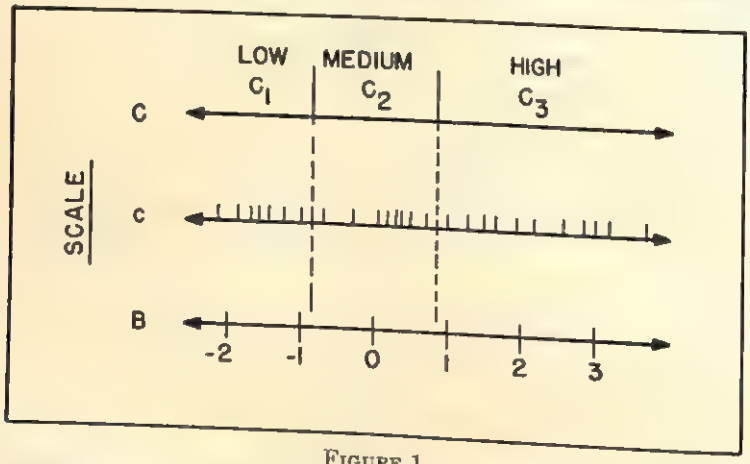


FIGURE 1

Conversion of a Sequence of Ordered Categories into an Interval Scale. Scale C Represents the Observed Rating Categories; Scale c, the Assumed Underlying Continuum; and Scale B, the Metric (with Arbitrary Origin and Unit Distance) Assigned to Scale c.

tinguishable esthetically. This is illustrated in Figure 1. Scale C is comprised of the categories C_1, C_2, C_3. Events falling in a category C_i are ordered as higher or lower than events in $C_j (i \neq j)$ for the property being rated, but no distinction is made among the events falling in C_i. Scale c is the continuum which is hypothesized or defined to underlie scale C, the smaller categories

indicating much finer differences in degree of the rated property. Strictly speaking, the underlying scale need not be an actual continuum; it may be conceptualized as a finite number of subdivisions of each category of the rating scale, as long as these subcategories are ordered by the same relation that orders the rating scale.

If numbers can now be assigned to the various positions on the underlying scale in such a manner that an interpretation can be defined, discovered, or assumed for the relative differences between positions on the scale, the theoretical underlying scale becomes an interval scale indicated in Figure 1 by scale B. Then the assignment of an event to a category of C is interpreted as an estimation of the score of the event on the underlying scale B. We shall refer to theoretical metrics such as B, defined or inferred from cruder empirical measures, as *base scales*. It has been cutomary to define the base scale (more rigorously, a set of base scales, linear transformations of one another) for a particular set of categories and distributions over the categories as that assignment of numbers to the continuum which normalizes the distributions (**10**). There may be other equally valid ways of defining the base scale, e.g., the counting of just-noticeable-differences, or defining the base scale so as to normalize distributions other than the ones being dealt with in the given study. It is not always possible, given more than one distribution over the same continuum, to find a numbering of the continuum that simultaneously normalizes all the distributions. Although the method of successive intervals, as described in the literature, has assumed normality for all distributions used in the analysis, we shall demonstrate that the validity of the method as a computational technique need not assume normal distributions.

Once the existence of a base scale has been defined over the categories of a rating scale, each event classified by the scale is considered to have a value on the base scale. Since each category corresponds to an interval of the base scale, the assignment of an event to a specific category determines a range in which its base scale value falls. If, now, the shape of a distribution of scores on the base scale is known, values for the widths of the various category intervals can be computed in terms of the standard deviation of that distribution as the unit of measurement. These are computed by tabulating the cumulative proportion of scores at each boundary of the interval and calculating the width in sigmas corresponding to such a percentile difference for that type of distribution. If the distribution is known or assumed to be normal, then the interval width will be the standard deviation of the distribution multiplied by the difference between the normal deviates corresponding to the cumulative proportions at the lower and upper boundaries of the interval. If a number of distributions are available, i.e., a group of judges rates a set of stimuli so that each stimulus determines a class of ratings, a number of measures of each interval width will be obtained in terms of the sigmas of the various distributions. If these are pooled,

estimates of the interval widths in terms of a common unit of measurement are obtained. Finally, if the median of a distribution be taken as the base scale score representing the distribution, the exact base scale value of this point can be estimated as follows: observe the cumulative proportion at the boundaries of the interval in which the median falls, compute from the assumed distribution function the proportion of the distance from the lower boundary, multiply this proportion by the interval width, and add the product to the base scale value for the lower boundary.

This and similar techniques for conversion of distributions of scores of a set of rating scale categories into points along an interval scale have been variously described in the literature, most frequently under the title, *the method of successive intervals* (**1**, **2**, **4**, **5**, **6**, **7**, **8**). The general computational steps usually given for evaluation of the interval widths are: (*i*) for each distribution, compute each interval width in terms of the sigma of that distribution by taking the difference between the normal deviates corresponding to the boundaries of the interval, assuming each distribution to be normal; (*ii*) let the average value of the computed widths for a given interval be taken as the best estimate of the width of that interval in terms of a unit of measurement common to all intervals. When the cumulative proportion at a boundary of an interval is nearly 0 or 1, the estimate of interval width given by that distribution for the interval is too unreliable for use, so the average width for an interval must be a weighted average; the weights of 0 or 1 have been employed in all past applications of the method.

Previous advocates of the method of successive intervals have attempted to validate the technique by demonstrating its extremely high correlation with the method of paired comparisions (**8**), and its internal consistency (**3**). It is our present aim to evaluate the method in terms of the degree to which results of the computations from empirical data can be expected to diverge from theoretically "true" values as determined from the definition of the base scale. That is, we propose to evaluate the *absolute* validity of the method.

The primary scores which are determined by the method of successive intervals are the widths of the category intervals relative to some arbitrary unit of measurement. The location of medians of the various distributions is secondary to estimation of the interval widths, since once the latter are known the former are easily determined. It is obvious that (*i*) if all the distributions used to measure the interval widths have equal variances, (*ii*) if all the distributions are normal, and (*iii*) if there are no sampling errors, then the computed values of relative interval widths are identical with the theoretical values. For, if each experimentally obtained distribution were normal, and for every distribution the proportion of cases falling within each interval showed no sampling errors, then the interval widths computed from a given distribution would be identical with the theoretical values as measured by the standard deviation of that distribution. If the variances

of all the distributions were equal, then each distribution would give the same computed value for a given interval width. Thus there are three possible sources of error in computation of base scale values by means of the method of successive intervals: type (*a*) errors due to unequal variances of the distributions used to compute the interval widths, type (*b*) errors due to non-normality of the distributions, and type (*c*) sampling errors, i.e., errors due to the estimation of cumulative proportions of the interval boundaries from finite samples of the measuring distributions.

As a tool for evaluation of the contributions of these sources of error to a total error of estimate of relative interval width, it is convenient to define a *coefficient of error*. Let a quantity χ be estimated by a quantity X. Then the coefficient of error, ξ, for the estimation of χ by X is $\xi = (X - \chi)/\chi$ or $X = (1 + \xi)\chi$. The magnitude of the coefficient of error gives the discrepancy between X and χ as a proportion of χ and is nothing more than 1/100 of the percentage error in the approximation of χ by X.

Relative interval widths computed by the method of successive intervals are estimates of true *relative* interval widths on the base scale. By relative interval widths, we recognize that the unit of measurement is arbitrary, so that the ratio of one interval width to another (which is invariant under transformation of the unit of measurement) is the critical quantity by which relative interval width is expressed. We can evaluate the coefficient of error for the estimation of true interval width ratios from computed ratios as follows: (*i*) find an expression for the computed interval widths, L_j and L_k, for categories j and k, in terms of the three types of errors that influence the computed widths, and of the true interval widths, λ_j and λ_k; (*ii*) set

$$L_j/L_k = (1 + \xi_{jk})(\lambda_j/\lambda_k). \tag{1}$$

Solving for ξ_{jk}, we obtain the coefficient of error for the estimation of relative interval widths by the method of successive intervals as a function of the different types of error; we shall be able to see explicitly the manner and extent to which each kind of error contributes to the total error.

Let λ_j be the true width on the base scale of an interval j in terms of some arbitrary unit of measurement U, and L_j the width of the interval as computed by the method of successive intervals. Let η_i (measured in terms of U) be the standard deviation of the ith distribution over the base scale. If the ith distribution is normal and displays no sampling errors, then the cumulative proportions at the upper and lower boundaries of interval j permit an exact computation (through use of a table of the normal proability integral) of the magnitude of λ_j in terms of η_i as a unit of measurement. Specifically,

$$l_{ij} = \lambda_j/\eta_i = \nu_i\lambda_j ,$$

where l_{ij} is the width of interval j as computed from distribution i; λ_j and η_i are the true magnitudes of interval j and the standard deviation of distri-

bution i, respectively, in terms of the arbitrary unit of measurement U; and $\nu_i = 1/\eta_i$. However, to the extent that distribution i diverges from normality and contains sampling errors, l_{ij} will differ from $\nu_i\lambda_j$. In general, $l_{ij} = \nu_i\lambda_j + \zeta_{ij}$, where ζ_{ij} is the discrepancy between the computed l_{ij} and the theoretical $\nu_i\lambda_j$. ζ_{ij} can be analyzed into two additive components E_{ij} and ϵ_{ij}, where E_{ij} is a *constant* bias due to non-normality of the distribution and ϵ_{ij} is a random sampling error. Thus,

$$l_{ij} = \nu_i\lambda_j + E_{ij} + \epsilon_{ij} . \tag{2}$$

It should be noted that the unit of measurement for the error terms E_{ij} and ϵ_{ij} is η_i, the standard deviation of the ith distribution; while the unit of measurement for $1/\nu_i$ and λ_j is the arbitrary U, which is the same for all distributions.

The computed width L_j of interval j is a weighted average of the estimates of widths contributed by the various distributions. That is,

$$L_j = \sum_i w_{ij}l_{ij} = \lambda_j \sum_i w_{ij}\nu_i + \sum_i w_{ij}E_{ij} + \sum_i w_{ij}\epsilon_{ij} , \tag{3}$$

where $\sum_i w_{ij} = 1$. Defining the quantities A_j, β_j, and γ_j by

$$A_j = \sum_i w_{ij}\nu_i , \tag{4}$$

$$\beta_j = (\sum_i w_{ij}E_{ij})/(A_j\lambda_j), \tag{5}$$

$$\gamma_j = (\sum_i w_{ij}\epsilon_{ij})/(A_j\lambda_j), \tag{6}$$

we obtain

$$L_j = A_j\lambda_j(1 + \beta_j + \gamma_j). \tag{7}$$

Since λ_j is inversely proportional to, and A_j proportional to the magnitude of the base unit of measurement U, $A_j\lambda_j$ is invariant for transformations of U. Therefore β_j and γ_j, which are also invariant under transformations of U, may be interpreted as error per unit length of interval due to non-normality of distribution and to sampling error, respectively. L_j may be interpreted as an estimate of λ_j with $1/A_j$ as the unit of measurement. It will be seen below that $1/A_j$ is approximately the harmonic mean of the standard deviations of the measuring distributions.

We are now able to evaluate the coefficient of error, ξ_{jk}, for the computed ratio L_j/L_k as an estimate of the true ratio λ_j/λ_k of the widths of intervals j and k. Finding L_k by substitution of k for j throughout (7) and solving for ξ_{jk} in (1) we find that

$$\xi_{jk} = \frac{A_j}{A_k}\left(\frac{1 + \beta_j + \gamma_j}{1 + \beta_k + \gamma_k}\right) - 1,$$

which may be written

$$\xi_{jk} = \alpha_{jk}\left(\frac{1 + \beta_j + \gamma_j}{1 + \beta_k + \gamma_k}\right) + \frac{\beta_j - \beta_k}{1 + \beta_k + \gamma_k} + \frac{\gamma_j - \gamma_k}{1 + \beta_k + \gamma_k}, \tag{8}$$

where

$$\alpha_{jk} = (A_j/A_k) - 1. \tag{9}$$

Since α_{jk} reflects the difference between units of measurement within interval j and within interval k, and vanishes (as shown below) when the variances of all the distributions are equal, α_{jk} may be regarded as the error in relative interval width due to unequal variances of the distributions which are used in estimating the interval widths. Thus, of the three sources of error in the method of successive intervals, type (*a*) is represented quantitatively by α, type (*b*) by β, and type (*c*) by γ.

γ-Error

It will be recalled from (3) that each distribution i was assigned a weight w_{ij} for its contribution l_{ij} in the computation of L_j. It is now possible to assign these weights in a manner that minimizes the sampling error γ_j. Assuming the various distributions to be essentially independent of one another in their sampling errors, we find from (6) the mean and variance for γ_j (under repeated sampling with a fixed set of weights) to be

$$\mu_{\gamma j} = \sum^{i} w_{ij}\mu_{\epsilon ij}/A_j\lambda_j\,; \tag{10}$$

$$\sigma^2_{\gamma j} = \sum^{i} w^2_{ij}\sigma^2_{\epsilon ij}/(A_j\lambda_j)^2. \tag{11}$$

But $\epsilon_{ij} = \delta_{Uij} - \delta_{Lij}$, where δ_{Uij} and δ_{Lij} are the sampling errors for the standardized deviates of the normal probability distribution corresponding to the cumulative proportions of distribution i at the upper and the lower boundaries of interval j, and $\delta \simeq \Delta P/y$, where ΔP is the sampling error of a cumulative proportion at an interval boundary and y is the ordinate of the normal probability distribution at P. Since the sampling mean of ΔP is zero, $\mu_{\epsilon ij} \simeq 0$ and hence, from (10),

$$\mu_{\gamma j} \simeq 0; \tag{12}$$

while

$$\begin{aligned}\sigma^2_{\epsilon ij} &= \sigma^2_{\delta Uij} + \sigma^2_{\delta Lij} - 2\operatorname{cov}(\delta_{Uij}, \delta_{Lij}) \\ &\simeq \frac{1}{n_i}\left[\frac{P_{Uij}(1 - P_{Uij})}{y^2_{Uij}} + \frac{P_{Lij}(1 - P_{Lij})}{y^2_{Lij}} - \frac{2P_{Lij}(1 - P_{Uij})}{y_{Lij}y_{Uij}}\right],\end{aligned} \tag{13}$$

where n_i is the sample size of distribution i, P_{Lij} (P_{Uij}) is the parametric cumulative proportion of distribution i falling at the lower (upper) boundary

of interval j, and $y_{L_{ij}}$ ($y_{U_{ij}}$) is the ordinate of the normal probability distribution at $P_{L_{ij}}$ ($P_{U_{ij}}$). Since the sample size n_i is known, and the sample cumulative proportions provide sufficiently close approximations to the parametric cumulative proportions, very close approximations of $\sigma^2_{\epsilon_{ij}}$ may be computed from empirical data by use of formula (13). Furthermore, $\sigma^2_{\epsilon_{ij}}$ may be made as small as desired by choosing the sample size n_i sufficiently large.

The assumption of (11) might at first seem gratuitous; for many experimental situations the sampling errors of one distribution will not be strictly independent of the next. Thus, if two sample distributions are obtained from judgments for two stimuli by the same judges, the sampling errors of the two distributions would probably be correlated. However, the disturbing effects of such a lack of strict independence are vitiated by the following considerations: (*i*) factors linking the sampling errors of two distributions usually comprise only a small portion of the total factors determining the outcome of the observed cumulative proportions; (*ii*) linear correlations among the sampling errors may be negligible even when significant non-linear correlations exist; and (*iii*) the intercorrelations may assume both positive and negative values, so that even when their absolute magnitudes are significant their net effect may be negligible. Thus the assumption of (11) involves little loss of generality.

Since by (12) the average value of γ_j is approximately zero, the expected absolute magnitude of γ_j is less than (though on the order of) σ_{γ_j}, so the expected (absolute) size of γ_j will be minimal when σ_{γ_j} is minimal. By differentiation of (11) it will be found that σ_{γ_j} is minimal when, for each i, $w_{ij}\sigma^2_{\epsilon_{ij}} = k_j$, where k_j is a constant of proportionality. Since $\sum_i w_{ij} = 1$, $k_j = (\sum_i \sigma^{-2}_{\epsilon_{ij}})^{-1}$, so

$$w_{ij} = (\sigma^2_{\epsilon_{ij}} \sum_i \sigma^{-2}_{\epsilon_{ij}})^{-1}. \tag{14}$$

Equations (14) and (13) provide the steps for computation of the proper weights. (Except for those distributions for which the cumulative proportion at one of the boundaries of an interval is close to 0 or 1, the weights assigned to the various distributions for that interval are very similar. Hence, the customary procedure of giving zero weight to those distributions for which the sampling reliability of the interval estimate is small and of giving the remaining distributions equal weight in the computation of the interval width should be acceptable for most purposes.)

Substitution of (14) in (11) gives

$$\sigma^2_{\gamma_j} = (A_j\lambda_j)^{-2}(\sum_i \sigma^{-2}_{\epsilon_{ij}})^{-1}. \tag{15}$$

Since $\sigma^2_{\epsilon_{ij}}$ is usually on the order of $1/n_i$, letting n be the average size of the sample distributions and N the number of distributions, $\sigma^2_{\gamma_j}$ is roughly on

the order of $(A_j\lambda_j)^{-2}$ $(nN)^{-1}$. Thus the expected order of magnitude for γ_j is roughly $(A_j\lambda_j)^{-1}$ $(nN)^{-\frac{1}{2}}$. This value may be made as small as desired by taking sufficiently large n and N. For example, if $A_j\lambda_j = .5$, $N = 50$ and $n = 500$, then the expected order of magnitude for γ_j is 10^{-2}. Thus for an empirical study of any substantial proportions an expected order of magnitude for γ_j of 10^{-2} should not be difficult to obtain. In order to maintain a fixed order of magnitude for γ_j, a decrease of interval widths must be compensated for by an increase in (*i*) the sample sizes, (*ii*) the number of distributions, or (*iii*) both. For with σ_{γ_j} held constant, $\sqrt{nN}$ is inversely proportional to $A_j\lambda_j$, while the latter, as shown below, is the width of category j in units of measurement given by the harmonic mean of the standard deviations of the measuring distributions. This has direct implications for the design of rating scales, for it shows that the number of categories into which a scale can be reliably decomposed is limited by the number of stimuli and the size of the population upon which the scale is to be standardized.

Thus, if the width of an interval relative to $1/A_j$ is not too small, and if the study by which the scale is being standardized is of reasonably substantial dimensions, the error in estimation of λ_j due to sampling will be insignificant—generally on the order of 10^{-2}. In light of this, the mean and sampling error of ξ_{jk} can be evaluated. Any reciprocal, $1/s_i$, from a distribution of s with mean M_s can be replaced by the expression $(2/M_s - s_i/M_s^2)$ with an error coefficient of $-[(s_i - M_s)/M_s]^2$. Since from (12), the mean of $(1 + \beta + \gamma)$ is $(1 + \beta)$, $[1/(1 + \beta_k + \gamma_k)]$ may be replaced by $[(1 + \beta_k - \gamma_k)/(1 + \beta_k)^2]$, with an error coefficient of $-[(\gamma_k)/(1 + \beta_k)]^2$, the error of the replacement being negligible so long as β_k does not approach -1. With this replacement we find from (8) and (12) that the sampling mean of ξ_{jk} is

$$\mu_{\xi_{jk}} \simeq \alpha_{jk}[(1 + \beta_j)/(1 + \beta_k)] + (\beta_j - \beta_k)/(1 + \beta_k), \tag{16}$$

while, disregarding second-order terms,

$$\sigma_{\xi_{ij}} \simeq [(1 + \alpha_{jk})/(1 + \beta_k)^2]\sqrt{(1 + \beta_k)^2\sigma_{\gamma_j}^2 + (1 + \beta_j)^2\sigma_{\gamma_k}^2}\,. \tag{17}$$

α-Error

For evaluation of α_{jk}, the error due to inequality of variances, it is convenient to employ the identity

$$A_j = \sum_{i=1}^{N} w_{ij}\nu_i = N\sigma_{w_j}\sigma_\nu r_{w_j\nu} + N\bar{w}_j\bar{\nu}$$

$$= N\sigma_{w_j}\sigma_\nu r_{w_j\nu} + \bar{\nu},$$

where N is the total number of distributions, σ_{w_j} is the standard deviation of the weights for the jth interval over the N distributions, σ_ν is the standard deviation of the ν_i over the N distributions, $r_{w_j\nu}$ is the product-moment

correlation between w_{ij} and ν_i over the N distributions, and $\bar{\nu}$ is the mean value of ν_i over the N distributions. From (9), this gives for α_{jk}

$$\begin{aligned}\alpha_{jk} &= \frac{N\sigma_{w_j}\sigma_\nu r_{w_j\nu} - N\sigma_{w_k}\sigma_\nu r_{w_k\nu}}{N\sigma_{w_j}\sigma_\nu r_{w_k\nu} + \bar{\nu}} \\ &= \left(\frac{C_j r_{w_j\nu} - C_k r_{w_k\nu}}{C_k r_{w_k\nu} V_\nu + 1}\right) V_\nu ,\end{aligned} \tag{18}$$

where $C_j = N\sigma_{w_j}$, $C_k = N\sigma_{w_k}$, and $V_\nu = \sigma_\nu/\bar{\nu}$.

The value of C_j depends only on the shape of the distribution of weights for the interval j; this value will be of an order higher than 10^{-1} only when a relatively small proportion of the distributions receive a significant weight for interval j. In the case where a proportion, k, of the distributions receive equal weights and the rest receive 0 weight, $C = \sqrt{(1/k) - 1}$, which exceeds 1 only when $k < .5$ and is no larger than 3 when $k = .1$. The correlation, $r_{w_j\nu}$, between the weights assigned to the distributions for interval j and the reciprocals of the standard deviations of the distributions can be expected to assume some small negative value (with a chance divergence which vanishes as N grows large), since as η_i increases the boundaries of the interval draw closer to the center of the distribution, yielding an increase in w_{ij}. However, we should expect this correlation to be equal for both intervals j and k. Thus, the maximum value of α_{jk} would be approximately $V_\nu \times 10^{-1}$.

But V_ν is the coefficient of variation for the reciprocals of the standard deviations of the measuring distributions and is approximately equal to the coefficient of variation for the η_i. We shall, as a rule, expect to find V_ν on the order of 10^{-1}, which makes α_{jk} on the order of 10^{-2}. Thus only when the variances of the distributions by which the interval widths are computed differ widely among themselves is the error contributed by the inequality of variances of any significance. In such cases, the data can be reanalyzed using the correction for inequalities in variance suggested by Attneave (1).

It should be noted that if all distributions receive equal weights for two intervals j and k, then $\alpha_{jk} = 0$, regardless of the magnitude of V_ν. Even when the η_i differ widely, α_{jk} will be negligible if the w_{ij} and w_{ik} are sufficiently homogeneous. It should also be noted that

$$A_j = (C_j r_{w_j\nu} V_\nu + 1)\bar{\nu} \simeq \bar{\nu}, \tag{19}$$

where $\bar{\nu}$ is the reciprocal of the harmonic mean of the η_i. This substantiates our earlier contention that the computed interval widths are expressed in units of measurement determined by the harmonic mean of the standard deviations of the measuring distributions.

β-Error

Of the three sources of error in the method of successive intervals, evaluation of β-error is the most difficult. We can replace $\sum_{i=1}^{N} w_{ij}E_{ij}$ by $(N\sigma_{w_j}$

$\sigma_{E_j} r_{w_j E_j} + \bar{E}_j)$, where $\bar{E}_j$ and σ_{E_j} are the mean and standard deviation of the errors introduced into the estimation of λ_j by the non-normality of the N distributions. Then from (5)

$$\beta_j = (C_j \sigma_{E_j} r_{w_j E_j} + \bar{E}_j)/(A_j \lambda_j). \tag{20}$$

Note that E_{ij} is measured in terms of the standard deviation, η_i, of distribution i. In particular, when $P_{L_{ij}}$ and $P_{U_{ij}}$ are the cumulative proportions of distribution i at the lower and upper boundaries of interval j, E_{ij} is the difference between the number of sigmas spanned between $P_{L_{ij}}$ and $P_{U_{ij}}$ by the actual distribution i and the number of sigmas spanned between $P_{L_{ij}}$ and $P_{U_{ij}}$ by a normal distribution. Let D_{ij} be the number of sigmas spanned by distribution i between these two cumulative proportions, and let d_{ij} be the corresponding number of sigmas spanned by a normal distribution. Then $E_{ij} = d_{ij} - D_{ij} = \omega_{ij} D_{ij}$, where $\omega_{ij} = (d_{ij}/D_{ij}) - 1$ and is thus the coefficient of error for the approximation of the distance in sigmas spanned between $P_{U_{ij}}$ and $P_{L_{ij}}$ by distribution i by the corresponding distance spanned by a normal distribution. Since $D_{ij} = \lambda_j/\eta_i = \nu_i \lambda_j$, (20) may be rewritten as

$$\begin{aligned}\beta_j &= (C_j r_{w_j E_j} \sigma_{\omega_j \nu \lambda_j} + \overline{\omega_j \nu \lambda_j})/(A_j \lambda_j) \\ &= C_j r_{w_j E_j} \sigma_{\omega_j (\nu/A_j)} + \overline{\omega_j (\nu/A_j)}.\end{aligned}$$

But when V_ν is small,

$$\sigma_{\omega_j (\nu/A_j)} \simeq \sigma_{\omega_j}$$

and

$$\overline{\omega_j (\nu/A_j)} \simeq \sigma_{\omega_j} V_\nu r_{w_j \nu} + \bar{\omega}_j .$$

Therefore,

$$\begin{aligned}\beta_j &\simeq \sigma_{\omega_j}(C_j r_{w_j E_j} + V_\nu r_{w_j \nu}) + \bar{\omega}_j \\ &\simeq \beta_j' + \bar{\omega}_j ,\end{aligned} \tag{21}$$

where

$$\beta_j' = \sigma_{\omega_j}(C_j r_{w_j E_j} + V_\nu r_{\omega_j \nu}). \tag{22}$$

In general, while there may be some small non-linear correlation between w_{ij} and E_{ij}, the linear $r_{w_j E_j}$ will be close to zero as N increases and the chance fluctuation of $V_{w_j E_j}$ thus diminishes. A similar argument holds for $r_{\omega_j \nu}$; because of the small expected values of C_j and V_ν, β_j' should be on the order of $\sigma_{\omega_j} \times 10^{-1}$ at maximum. It will be shown below that even when a distribution is markedly non-normal, the expected order of magnitude for ω_{ij} is only 10^{-1}, so σ_{ω_j} will be on the order of 10^{-1} at maximum. Thus, β_j' will be on the order of 10^{-2} at maximum and is more likely to be of order 10^{-3}.

It follows that the only likely significant component of β_j is $\bar{\omega}_j$, the latter comprising the average value of ω_{ij} over the N distributions used to measure the width of interval j. These N distributions may be conceived as a sample of size N from an infinite population of potential distributions over the scale. Then $\bar{\omega}_j$ has a sampling mean, $\mu_{\bar{\omega}j}$, and variance, $\sigma^2_{\bar{\omega}j}$, of its own. Similarly, the ω_{ij} for the infinite potential population of distributions over interval j have a mean, μ_{ω_j}, and variance, $\sigma^2_{\omega_j}$. Finally, since $\bar{\omega}_j$ is the mean of a sample of size N from the ω_{ij}, $\mu_{\bar{\omega}j} = \mu_{\omega_j}$ and $\sigma^2_{\bar{\omega}j} \simeq \sigma^2_{\omega_j}/N_j$. (More generally, $\sigma^2_{\omega_j}/N \leq \sigma^2_{\bar{\omega}j} \leq \sigma^2_{\omega_j}$, depending upon the extent to which the ω_{ij} for the sample of N distributions are independent of one another. In most situations, we will expect to find that the ω_{ij} are not wholly independent, but, for the same reasons advanced to justify equation (11), we shall expect that the sum of the covariances will be negligible.) Let β_j'' be the extent to which $\bar{\omega}_j$ diverges from its mean. Then

$$\bar{\omega}_j = \beta_j'' + \mu_{\omega_j}, \tag{23}$$

and thus, from (21),

$$\beta_j = \beta_j' + \beta_j'' + \mu_{\omega_j}. \tag{24}$$

Since β_j'' is of order $\sigma_{\omega_j}/\sqrt{N}$, and as already mentioned, the expected order of magnitude for σ_{ω_j} is 10^{-1} or smaller, then if N is reasonably large the maximum expected order of magnitude for β_j'' is 10^{-2}. This leaves μ_{ω_j} in (21) as the only component of β_j likely to be significant. But μ_{ω_j} is merely the expected value of ω_{ij} on the interval j. As indicated below, the absolute magnitude of ω_{ij} is only of expected order 10^{-1} even when the distributions are quite non-normal. While it is impossible to make any definite statement about the average, μ_{ω_j}, for an interval j, it would seem unlikely that it could exceed .10 except under cases of extreme, persistent, and positively correlated non-normalities among the population of distributions over interval j. Thus, except under unusual circumstances, β_j is of expected order of magnitude 10^{-1} or less, and we may simplify (16) and (17) to

$$\begin{aligned} \mu_{\xi_{jk}} &\simeq \alpha_{jk} + \beta_j - \beta_k \\ &\simeq \alpha_{jk} + (\beta_j' - \beta_k') + (\beta_j'' - \beta_k'') + \mu_{\omega_j} - \mu_{\omega_k} \end{aligned} \tag{25}$$

and

$$\sigma_{\xi_{jk}} \simeq \sqrt{\sigma^2_{\gamma_j} + \sigma^2_{\gamma_k}}. \tag{26}$$

Of the terms in (25), only μ_{ω_j} and μ_{ω_k} are of expected order larger than 10^{-2}.

It yet remains to determine the anticipated order of magnitude for ω. Since the population of potential distributions over a rating scale cannot be specified, it is impossible to assign a mathematical expectation to this term. However, ω may be computed as a function of the degree of non-normality of the distribution being approximated. One may then select a range of distributions within which an empirically encountered distribution reasonably may be anticipated to fall and hence obtain reasonable bounds

for the magnitude of ω. What we shall illustrate here is a technique by which a distribution of any given shape readily may be inspected for its values of ω. By this technique, the reader may select what he considers to be fair examples of empirically anticipated non-normal distributions and easily convince himself that ω is unlikely to be of an order greater than 10^{-1}.

It will be recalled that $\omega = (d/D) - 1$, where d and D are the distances, measured in terms of the standard deviations of the distributions, spanned between the cumulative frequencies at the upper and lower boundaries of the interval by a normal distribution and the empirical distribution, respectively. Let P_U and P_L be the cumulative proportions at the upper and lower boundaries of the interval, let $y(x)$ and $Y(x)$ be the ordinates at x of the unit normal distribution and of the empirical distribution standardized to $\sigma = 1$, respectively, and let x_P and X_P be the distance of cumulative proportion P from the means of the unit normal distribution and the standardized empirical distribution, respectively. Then $d = x_{P_U} - x_{P_L}$ and $D = X_{P_U} - X_{P_L}$. But

$$P_U - P_L = \int_{x_{P_L}}^{x_{P_U}} y(x)\, dx = \bar{y}d,$$

where $\bar{y}$ is the mean value of the ordinate to the unit normal distribution over the interval. Similarly,

$$P_U - P_L = \int_{X_{P_L}}^{X_{P_U}} Y(x)\, dx = \bar{Y}D.$$

Hence $d/D = \bar{Y}/\bar{y}$, so $\omega = (\bar{Y}/\bar{y}) - 1$ and is thus the coefficient of error for the approximation of the average height of the unit normal distribution between two cumulative proportions by the corresponding average height of the standardized empirical distribution. The magnitude of ω is then readily seen by an inspection of the graphs of y and Y against P. That is, let $y(P) = y(x_P)$ and $Y(P) = Y(X_P)$. It is computed without difficulty that $\bar{y}(x) = \bar{y}(P)$, where $\bar{y}(P)$ is the harmonic mean of $y(P)$ between P_L and P_U, and similarly $\bar{Y}(X) = \bar{Y}(P)$. Also, except for those intervals over which the coefficient of variation for $y(P)$ or $Y(P)$ is large, $\bar{y}(P) \simeq \bar{y}(P)$ and $\bar{Y}(P) \simeq \bar{Y}(P)$. Thus, given any empirical distribution, the magnitude of the approximation error can be determined readily by standardizing the distribution to unit variance, graphing the height of the distribution against cumulative proportion, and superimposing the corresponding graph of the unit normal distribution. One may then select two cumulative proportions, estimate the average difference between the curves over the interval visually, and divide this by the estimated average ordinate of the normal distribution over the interval.

We illustrate the method through its application to two arbitrary distributions, a rectangular distribution and a triangular distribution skewed

so that the projection of the apex divides the base in a ratio of 1:3. These are shown with unit variance in Figure 2, together with the unit normal distribution by which they are to be approximated. Both distributions represent departures from the normal that, in an empirical distribution, would be considered severe. Figure 3 shows the same distributions in terms

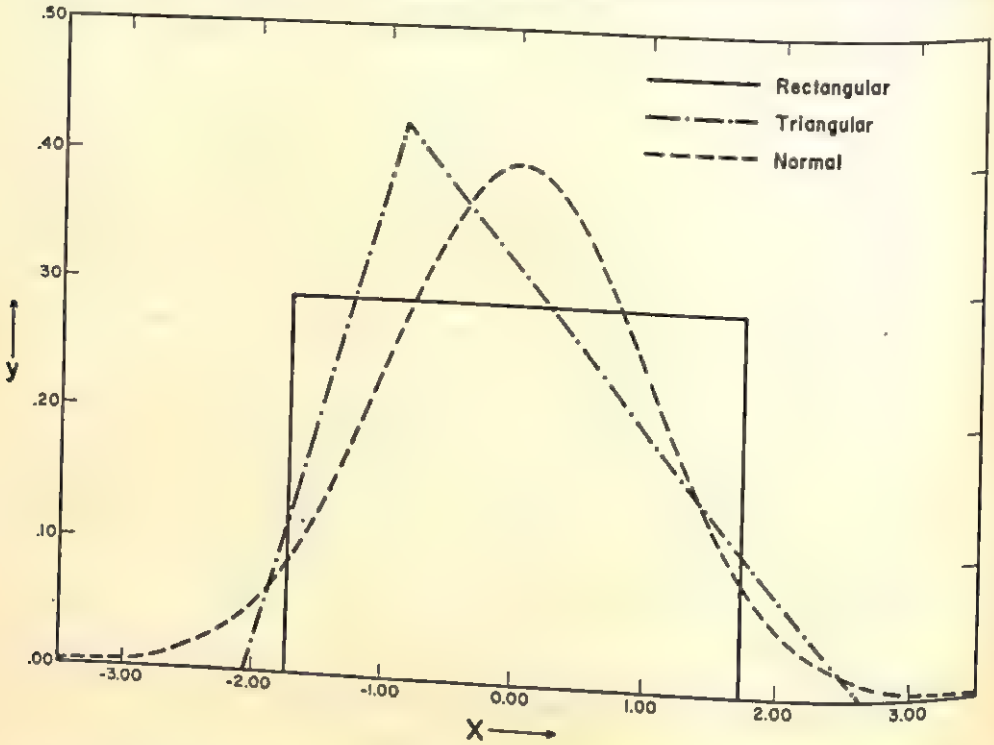


FIGURE 2
Normal, Rectangular, and Triangular Probability Distributions for Which $\mu = 0$, $\sigma = 1$. Projection of Triangular Apex Divides the Base into 1:3 Ratio.

of the ordinates of Figure 2 plotted against the corresponding cumulative proportions. If various pairs of cumulative proportions are selected and ω estimated, it is seen that $|\omega|$ has a modal value in the range.25 to .30 for the two distributions and grows much larger than this only when one of the proportions approaches 0 or 1 (due, here, to the finite ranges of both illustrative distributions). This is typical for most distributions; ω is likely to exceed the order of 10^{-1} only when one of the cumulative proportions at an interval boundary approaches the upper or lower limit. But it is precisely in this case that the error variance of a proportion obtained through finite sampling becomes so large as to give negligible weight to the contribution to the total estimate by an interval width estimate based on such a proportion.

In those few cases where an empirical distribution is likely to show large approximation errors (such as the case of multimodal distribution in which

the modes are well separated and the intervening troughs deep) the severe non-normality of the distribution should be painfully apparent when the distribution is plotted on the successive intervals scale as finally computed. The non-normal distribution then may be discarded and a new analysis of the remaining data performed if the investigator sees fit.

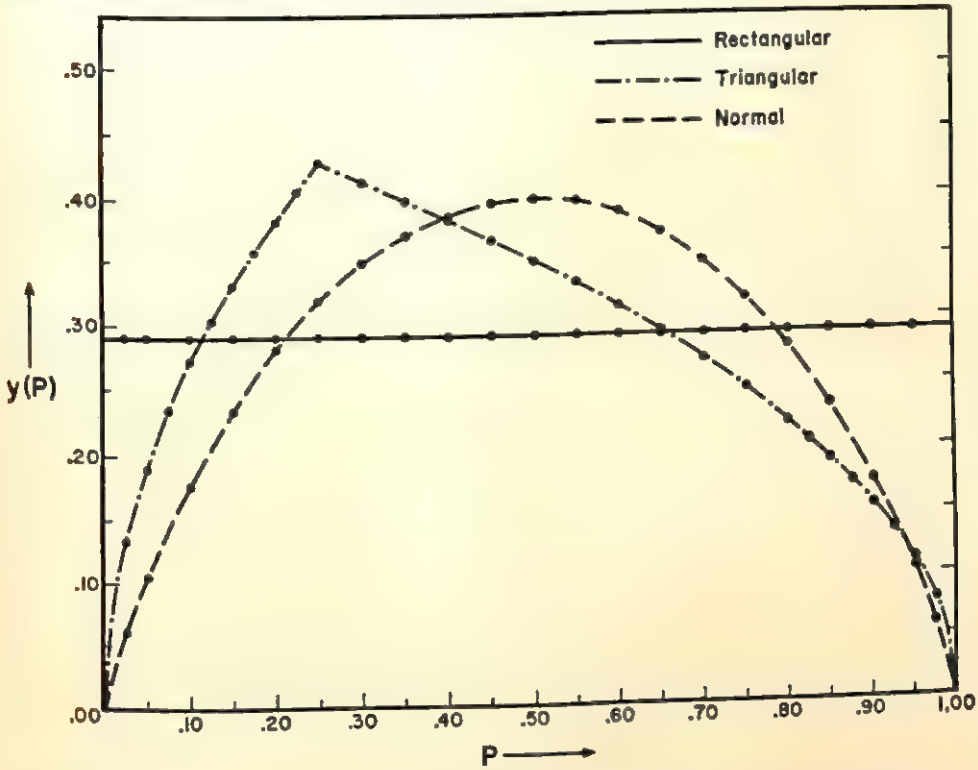


FIGURE 3
Ordinates of the Distributions of Figure 2 as Function of Cumulative Proportions.

β-Error and Its Relation to the Base Scale

So far we have found it unnecessary to make any comments concerning the base scale which supposedly underlies the rating scale except to hypothesize its existence.

Necessary and sufficient conditions for the existence of an *interval* base scale underlying a successive interval rating scale are: (*i*) There must (potentially) exist a numbering of all the potentially infinite number of events classifiable by the rating scale such that there is no overlap, for any two categories of the rating scale, of the ranges of the numbers corresponding to the events falling within each category. (*ii*) The positions of the ranges corresponding to the various categories must be in the same ordinal relation as are the categories. (*iii*) For any set of events so numbered, there must exist some interpretation of (*a*) the ordinal relations among the numbers

assigned to members of the set and of (*b*) the ratio of the difference between the numbers of any pair of the set to the difference between the numbers of any other pair, in terms of some properties of the events of the set. (If the base scale provides interpretations for additional properties of the numbers assigned to events, it may become a *ratio* or even an *absolute* scale.) There may be many different numberings satisfying conditions (*i*) and (*ii*), and many different interpretations in accordance with condition (*iii*). Hence there may be many different base scales underlying a given successive intervals scale. In fact, any assignment of numbers satisfying conditions (*i*) and (*ii*) is a potential base scale for the rating scale since we can never know for certain that there exists no interpretation of a numbering in conformance with condition (*iii*). Such potential base scales for a given rating scale need be correlated only to the extent that the values of a given event on the various potential base scales must all fall within ranges corresponding to the same rating scale category. In particular, any order-preserving transformation of a potential base scale is also a potential base scale.

The essential result of a method of successive intervals analysis is the derivation of a set of numbers corresponding to the boundaries of the intervals of the rating scale; these numbers, when paired and the ratio of differences between members of pairs taken, give the ratio of the base scale intervals corresponding to these pairs. The ratio of intervals for a potential base scale is always the same as the corresponding ratio for any *linear* transformation of that scale. However, this is not uniformly true for any other transformation. Let all potential base scales be separated into classes, any member of a given class being a linear transformation of any other member of that class. These classes are, in general, characterized by different values for the ratio of two intervals corresponding to two pairs of points on the rating scale; the classes of potential base scales most closely approximated by the scale computed through the successive intervals technique will be those classes whose ratios for interval widths are most similar to the ratios displayed by the computed scale. That is, the classes of potential base scales most closely approximated by the computed scale are the classes of potential base scales minimizing the ξ_{jk}.

Since the exact magnitudes of the ξ_{jk} are unknown in applications of the method of successive intervals, it is impossible to determine the class of potential base scales most closely approximated in a specific instance. The classes of potential base scales *most likely* to minimize the ξ_{jk}, however, are those which minimize the *expected* values of the ξ_{jk}. Now, the data of a specific successive intervals analysis are obtained by sampling of two kinds: a sample of size N from possible distributions over the rating scale, and a sample of n_i individuals from each distribution i ($i = 1, 2, \cdots, N$). The expected value of ξ_{jk} for a specific sample of distributions is given by (25). But the terms α_{ik}, $(\beta_j' - \beta_k')$, β_j'', β_k'' are dependent upon the specific sample of

distributions chosen; β_j'' and β_k'', by definition, have an expected value of 0, while both α_{jk} and $(\beta_j' - \beta_k')$ are determined essentially by differences of the form $X_j r_{y_j z_j} - X_k r_{y_k z_k}$, where X, y, and z are various specified properties of the N distributions. These differences should be negative as often as positive, so that the expected values of α_{jk} and $(\beta_j' - \beta_k')$ should be 0. Thus, the expected value of ξ_{jk} is approximately $\mu_{\omega_j} - \mu_{\omega_k}$, and hence the classes of potential base scales most closely approximated by the expected computed scale are those base scales showing the smallest differences among the μ_{ω_j}, μ_{ω_k}, $\cdots$ for the various intervals j, k, $\cdots$ of the scale. This important conclusion may be rephrased as: *the classes of potential base scales expected to be most closely approximated by the method of successive intervals are the classes for which the average coefficient of error (for the estimation of interval widths under assumptions of normality) is most nearly the same for all intervals of the rating scale.*

In particular, if, as implicitly assumed by previous psychometric analyses wherein the base scale remained unidentified, there exists a class of potential base scales which simultaneously normalize all distributions, then $\mu_\omega = 0$ for all intervals of these scales; there is no class of base scales more closely approximated by the expected computed scale.

Thus, we see that there is no single answer to the question of the magnitude of error involved in the approximation of an unidentified base scale by the method of successive intervals; the magnitude of error is relative to that base scale for which the computed scale is considered an approximation. If we wish, however, we may define the base scale to be approximated as that scale which simultaneously equalizes the μ_ω for all intervals. A class of such scales can always be found, and further, the set of all such classes includes all base scales which simultaneously normalize all distributions over the rating scale if such scales exist. If the base scale is so defined, then from (25)

$$\mu_{\xi_{jk}} = \alpha_{jk} + (\beta_j' - \beta_k') + (\beta_j'' - \beta_k''). \tag{27}$$

Only when the measuring distributions are extraordinarily non-normal are any of the terms on the right side of (27) of expected magnitude greater than 10^{-2}, and thus ξ_{jk} has an expected order of magnitude of no greater than 10^{-2}. This, in conjunction with (26), shows that if the sample sizes of the distributions have been taken sufficiently large (say, large enough to make σ_γ on the order of 10^{-2}), then *the extent to which interval ratios computed by the method of successive intervals diverge from the corresponding theoretically "true" values should not exceed* 10 *per cent of the latter, and may be much smaller if the experimental study has been well designed.*

Conclusions

Abstracting the essentials of the foregoing analysis, three major points are of significance—the first, a contribution to the computation technique of the method of successive intervals; the second, an evaluation of the validity

of the method; and the third, the significance of the method for the basic methodology of psychophysical measurement.

The contribution to computational technique is given by equations (13) and (14); it involves the computation of weights for the estimates of a given interval width so as to minimize the sampling errors for the composite estimate of the interval width. Except for the more exacting studies, however, or unless suitable tables have been obtained, the improvement of this exact method of weighing over the more rough and ready techniques now in use will scarcely be worth the extra computational labor. Of greater potential application in the design of empirical studies is the determination of the, relations among width of interval, the number of measuring distributions and their sample sizes for the maintenance of a fixed level of freedom from sampling error.

The validity and reliability of the method of successive intervals do not depend upon normality of distributions or equality of their variances. The reliability, as attested by (26), may be made as high as desired. If the base scale is suitably defined (i.e., defined so as to equalize, for the various intervals, the error due to estimation of interval width from a table of the normal probability integral) and if the reliability is made sufficiently high, then the validity, as implied by (27), is so high as to lead to an expected coefficient of error for relative interval widths of no more than a few parts in a hundred. Further, this validity is in reference to the theoretical values of the interval ratios. It is thus an *absolute* validity in contrast to past validation of psychophysical scaling techniques, where validation is attempted only in terms of internal consistency or consistency among different techniques purported to compute the same base scale. It would appear, then, that until similar analyses can be constructed for other psychophysical scaling techniques, the method of successive intervals should be accepted as the basic standard against which other techniques are to be validated.

Finally, and probably most important of all, we consider the implications of this analysis for the methodology of psychophysical measurement. It has been shown that it is unnecessary for psychophysical measurement (or for that matter, for any form of measurement) to assume any specific form of distributions over a measuring scale. The only assumption required is that certain properties of the measurements obtained by the measuring technique have some potential interpretative significance. The major premise of psychometric scaling in the past has been that if (*a*) a scale can be obtained which normalizes the distributions over it, then (*b*) that scale, or another very similar to it, has interpretive significance as an interval scale. We may now replace this premise with another: if (*a'*) a scale can be obtained which equalizes, for all intervals, the average coefficient of error for the approximation of interval width by the distance which normal distributions of equal standard deviations would span between corresponding percentiles,

then (*b*) that scale, or another very similar to it, has interpretive significance as an interval scale. The latter premise is both weaker and stronger than the former: weaker in that a scale satisfying (*a'*) can always be found, and such a scale also satisfies (*a*) when scales satisfying (*a*) exist; stronger in that the latter premise demands a meaningful scale to underlie every psychophysical measuring technique, whereas the former demands such a meaningful understructure only if a psychometric scale can be found to normalize simultaneously all distributions over it. Actually, the (*b*) clause of these premises is not so strong as it might appear. In a certain sense, the mere act of defining a scale in terms of the distributions over it imparts a meaning to the scale values so defined. Essentially, what our present analysis has shown is that it is always possible to give a distributional definition to a base scale which simultaneously normalizes all distributions regardless of whether or not a scale exists.

Since interpretation of psychometric scales has been sought in actual practice, regardless of whether simultaneous normalization could be realized, it is essential, if psychometric custom now current is to be justified, that a way be found to define psychometric scales in terms of properties other than such normalization. It is our belief that such justification has now been furnished.

REFERENCES

1. Attneave, F. A. A method of graded dichotomies for the scaling of judgments. *Psychol. Rev.*, 1949, **56**, 334–340.
2. Edwards, A. L. The scaling of stimuli by the method of successive intervals. *J. appl. Psychol.*, 1952, **36**, 118–122.
3. Edwards, A. L. and Thurstone, L. L. An internal consistency check for scale values determined by the method of successive intervals. *Psychometrika*, 1952, **17**, 169–180.
4. Garner, W. R. and Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.
5. Guilford, J. P. The computation of psychological values from judgments in absolute categories. *J. exp. Psychol.*, 1938, **22**, 32–42.
6. Guilford, J. P. Psychometric methods, 2nd ed. New York: McGraw-Hill, 1954 (pp. 223–262).
7. Mosier, C. I. A modification of the method of successive intervals. *Psychometrika*, 1940, **6**, 101–107.
8. Saffir, M. A. A comparative study of scales constructed by three psychophysical methods. *Psychometrika*, 1937, **2**, 179–198.
9. Stevens, S. S. Mathematics, measurements, and psychophysics. In S. S. Stevens (Ed.): Handbook of experimental psychology. New York: Wiley, 1951.
10. Thurstone, L. L. A law of comparative judgment. *Psychol. Rev.*, 1927, **34**, 273–286.

Manuscript received 3/10/54

Revised manuscript received 3/18/55

PSYCHOMETRIKA—VOL. 21, NO. 2
JUNE, 1956

RELATIONSHIPS BETWEEN TWO SYSTEMS OF FACTOR ANALYSIS

C. W. HARRIS

UNIVERSITY OF WISCONSIN

Considering only population values, it is shown that the complete set of factors of a correlation matrix with units in the diagonal cells may be transformed into the factors derived by factoring these correlations with communalities in the diagonal cells. When the correlations are regarded as observed values, the common factors derived as a transformation of the complete set of factors of the correlation matrix with units in the diagonal cells satisfy Lawley's requirement for a maximum likelihood solution and are a first approximation to Rao's canonical factors.

One of the distinctions in factor analysis that may be made, when viewed at the procedural level, is the distinction between choosing to factor R_1, the matrix of the intercorrelations of the variables with units in the diagonal cells, and choosing to factor R, the matrix with communalities in the diagonal cells. The purpose of this paper is to develop the transformation which relates the factors F_1 of R_1 to the factors F_2 of R. In order to develop this relationship, it will be assumed first that the elements of R_1 and of R have been determined without error, i.e., that the correlations and the communalities are *population* values. Guttman (2) has discussed conditions that are necessary for common-factor, or communality, solutions. Considering only population values, this paper shows that if a communality solution exists, it is simply a projective transformation of the set of complete factors of R_1.

The problem of estimation will next be considered. It will be shown that the transformation developed here when applied to *observed* data yields common factors that satisfy Lawley's (5, 6) requirements for a maximum likelihood solution and are a first approximation to factors derived by Rao's (7) canonical factor analysis. In developing these points, certain simplifications have been introduced for convenience. For example, only orthogonal factor solutions are employed; however, such solutions may be rotated to oblique factors if such are desired. Second, the factoring of variances and covariances, rather than R_1, is considered only incidentally in connection with Rao's canonical factor analysis. Third, it is generally assumed that R_1, and consequently F_1, is nonsingular. This is a realistic assumption. If necessary it can be abandoned, provided we make certain modifications in the algebra; these modifications do not negate the generalizations given here.

The distinction between choosing to factor R_1 and choosing to factor R is of considerable importance theoretically. In the former case the factors

are conceived to be located within the space defined by the variables viewed as vectors, whereas in the latter case both the common and the unique factors are conceived to be located in a space distinct from the space defined by the variables viewed as vectors. Consequently, factoring R_1 is equivalent to developing both the factor coefficients and the factor scores by appropriate linear operations on Z, where $ZZ' = R_1$; whereas the communality principle implies that a projection of Z, say ZA, is the matrix that is operated on to develop both the factor coefficients and the factor scores. Cf. (3). Since A is unknown, it follows that for the communality principle the factor scores cannot be computed, even though the elements of R are determined without error, but must be estimated.

It is well known (1, 4) that we may define an arbitrary orthogonal factor solution in terms of linear operations on a rectangular matrix of order n by N. Let Y be any n by N matrix, $n < N$. Then we may resolve Y into a product of two matrices, $Y = FS$, such that SS' is an identity matrix of the appropriate order. It is conventional to call F the factor matrix and S the matrix of factor scores. Various factoring methods, such as the centroid, the diagonal, or the principal-axis methods, can be identified in terms of rules for choosing the linear operators.

Now let us examine the relationship between factors extracted from R_1 and those extracted from R. If both Z and ZA are given, we may write

$$Z = F_1 S_1 , \tag{1}$$

by which is implied that the matrix Z has been factored completely, usually with as many factors as variables; and

$$ZA = F_2 S_2 , \tag{2}$$

with the implication that ZA has been factored completely, yielding $m \leq n$ factors. It has been shown (3) that $S_2'S_2$ defines A, which is the symmetric idempotent matrix that achieves the projection of Z that is required by the communality principle. Therefore, since $S_2S_2' = I$ of order m, we may write

$$ZS_2'S_2 = F_2 S_2 ,$$

and

$$ZS_2' = F_2 .$$

Then the desired relationship is given by

$$F_2 = F_1(S_1 S_2'). \tag{3}$$

The matrix in parentheses in (3) will ordinarily be singular; if so, it is not possible to solve (3) to write F_1 as a transformation of F_2 .

The matrix in parentheses in (3) is the transformation that relates the factors derived by the communality principle to those derived from the complete factoring of R_1 . That this is the relationship to be expected on other grounds is readily seen. Any matrix of unit-length factor scores, S, may be viewed as a matrix of direction cosines that gives the location of the factors with respect to the N person vectors, i.e., they locate the factors in the person space. With two sets of orthogonal factors located in the same person space, the transformation of the inner products of the variables with one set of factors to the inner products of the variables with the other set of factors is given by the correlations between the two sets of factors. It should be noted, however, that this is not a conventional orthogonal rotation; for example, the sums of squares of the entries in any column of the transformation need not be unity.

A solution for the transformation $T = (S_1S_2')$ may be obtained. For example,

$$(F_1'F_1)^{-1}F_1'F_2 = T$$

follows from (3). This is a true equation and not merely a least squares approximation under the conditions described here. This is to say that the symmetric idempotent matrix generated from F_1 is a unit for multiplication of F_2 . If R_1 is nonsingular, then F_1 also is nonsingular, and so the solution for T might be written

$$(F_1)^{-1}F_2 = T.$$

Another means of solving for T follows from the requirement that

$$F_2F_2' = R = F_1TT'F_1' = R_1 - U^2 = F_1F_1' - U^2,$$

where U^2 designates the matrix of unique variances. Assuming R_1 nonsingular,

$$TT' = I - (F_1)^{-1}U^2(F_1')^{-1}. \tag{4}$$

A solution for T (or some orthogonal rotation of T) is then given by factoring the matrix on the right of (4). At this point a solution for T is merely of theoretical interest, since if population values of correlations and communalities were known, the straightforward approach to determining the common factors would be to factor R.

Estimation of Common Factors

An important statistical problem faced by the factor analyst who wishes to employ the communality principle is that of estimation. Lawley (5, 6) has presented a maximum likelihood solution that has certain optimum characteristics from the statistical point of view. He also presents a test of significance for common factors derived by his iterative procedure. Rippe

(8) has extended the test of significance to factors derived by any method. Lawley's requirement on the factors F_2 is, in the notation employed here,

$$DF_2' = F_2'(U^{-2}R_1 - I), \tag{5}$$

where D is a diagonal matrix. Regard $R_1 = F_1F_1'$ as observed values. Lawley's development shows that given m common factors with U^2 as estimates of the unique variances of the variables, values of F_2 that satisfy (5) generate a reproduced matrix R^* such that a test of significance of the residual matrix $R_1 - R^*$ may be made.

Modify (5) by substituting $F_1T = F_2$, and $R_1 = F_1F_1'$; then with F_1 non-singular,

$$DT' = T'(F_1'U^{-2}F_1 - I). \tag{6}$$

This states that to satisfy Lawley's requirement, each column of T must be proportional to a unit-length characteristic vector of the matrix $(F_1'U^{-2}F_1 - I)$. From (4) it is evident that a satisfactory solution for T is to choose each column as proportional to a unit-length characteristic vector of the matrix on the right of (4), i.e., to define T by a principal-axis factoring of this matrix. But this is merely restating the requirement of (6), since the characteristic vectors of $(F_1'U^{-2}F_1 - I)$ and of $[I - (F_1)^{-1}U^2(F_1')^{-1}]$ necessarily are the same. We have therefore shown that the solution $F_2 = F_1T$ satisfies Lawley's requirement when T is defined by a principal-axis factoring of the matrix on the right of (4).

Now it is at least intuitively evident that one or more columns of T might be null or consist of imaginary numbers. Let us adopt a rule that refuses to admit such columns in any solution of F_2. This implies that the number of common factors must be considered, as well as the estimates of unique variance. If at least one column of T is admissible, then the resulting factor or factors generate a reproduced matrix such that the residuals may be tested for significance by Lawley's procedure. If the residuals are not significant, then an upper bound to the number of common factors has been determined. If they are significant, new trial values of U^2 must be employed, and a new determination of the number of admissible columns of T and of the significance of the residuals must be made.

A relatively new attack on the problem of estimation of common factors is given by Rao's (7) canonical factor analysis, which is one of many possible maximum likelihood solutions. Rao derives his basis of estimation by requiring the correlation between a linear combination of ZA and a linear combination of Z to be a maximum. He calls this canonical factor analysis because of its connection with canonical correlation theory. His estimation procedure modifies initial trial values of U^2 by an iteration process, under the restriction of a given number of common factors. A test of significance for a least number

of factors is provided. Rao also shows that canonical factor loadings derived from correlations are proportional to those derived from covariances, with the constants of proportionality given by the sample standard deviations of the variables. Therefore, in showing that the solution $F_1T = F_2$ is a first approximation to Rao's canonical factors, we imply that the solution based on variances and covariances, instead of R_1, also is a first approximation to canonical factors.

Rao's procedure requires that we select the non-zero elements of U^2 to satisfy

$$1/u_i^2 = (\lambda_1 - 1)a_{i1}^2 + (\lambda_2 - 1)a_{i2}^2 + \cdots + (\lambda_m - 1)a_{im}^2 + 1, \tag{7}$$

where each λ_j is a latent root of the matrix $U^{-1}F_1F_1'U^{-1}$, and each a_{ij} the appropriate element of the unit-length characteristic vectors of the same matrix. Now these roots and vectors are connected with the roots and vectors of TT' in a regular manner. This connection is derived from the fact that the roots of $U^{-1}RU^{-1} = (U^{-1}R_1U^{-1} - I)$ are all one less than the roots of $U^{-1}R_1U^{-1}$, and the characteristic vectors of $U^{-1}RU^{-1}$ and $U^{-1}R_1U^{-1}$ are identical. Let L^2 designate the roots of $U^{-1}F_1F_1'U^{-1}$; then the roots of TT' are $(I - L^{-2})$. Conversely, let the positive roots of TT' be designated by D_m^2. Then $(I - D_m^2)^{-1}$ yields the m roots in L^2 that are each greater than unity; call this matrix L_m^2. Let Q designate the unit-length rows of characteristic vectors of $U^{-1}F_1F_1'U^{-1}$. Then the m unit-length rows of characteristic vectors of TT' coresponding to the positive roots D_m^2 are given by $L_m^{-1}QU^{-1}F_1$. It is now evident that $P'D_m$, defining m admissible columns of T, yields by the above transformations the required values for substitution in (7). Apparently, then, for any specified U^2 we may characterize the solution $F_1T = F_2$ as a first approximation to Rao's canonical factors. This is verified by noting that Rao defines canonical factors, at any stage of approximation, by

$$F_2 = UQ'(L_m^2 - I)^{1/2}. \tag{8}$$

Substituting for P' and D_m their expressions in terms of roots and vectors of $U^{-1}F_1F_1'U^{-1}$ and then simplifying gives

$$F_1T = F_1P'D_m = UQ'(L_m^2 - I)^{1/2},$$

which is identical with (8).

This demonstration of the connection of T with Rao's procedure is best characterized as making explicit an alternate path to canonical factors. It seems practically certain that Rao recognized the existence of this alternate path and rejected it for a very practical reason, namely, that his calculation routine is less laborious than one based on finding F_2 by way of F_1.

REFERENCES

1. Guttman, Louis. General theory and methods for matrix factoring. *Psychometrika*, 1944, 9, 1–16.
2. Guttman, Louis. Some necessary conditions for common-factor analysis. *Psychometrika*, 1954, 19, 149–61.
3. Harris, Chester W. Separation of data as a principle in factor analysis. *Psychometrika*, 1955, 20, 23–28.
4. Holzinger, Karl J. Factoring test scores and implications for the method of averages. *Psychometrika*, 1944, 9, 155–67.
5. Lawley, D. N. The estimation of factor loadings by the method of maximum likelihood. *Proc. Royal Soc. Edinburgh*, 1940, 60, 64–82.
6. Lawley, D. N. Further investigations in factor estimation. *Proc. Royal Soc. Edinburgh*, 1941, 61, 176–85.
7. Rao, C. R. Estimation and tests of significance in factor analysis. *Psychometrika*, 1955, 20, 93–111.
8. Rippe, Dayle D. Application of a large sampling criterion to some sampling problems in factor analysis. *Psychometrika*, 1953, 18, 191–205.

Manuscript received 10/6/54

Revised manuscript received 4/4/55

PSYCHOMETRIKA—VOL. 21, NO. 2
JUNE, 1956

TESTING MESSAGE DIFFUSION IN HARMONIC LOGISTIC CURVES*

STUART CARTER DODD
UNIVERSITY OF WASHINGTON

The growth of a population of knowers of a message was studied to test a human interactance hypothesis. The conditions investigated involved *people interacting in time*, with the population pairing off randomly (i.e., determined by many, small, different influences) and transferring an attribute (i.e., an all-or-none act) at either a steady rate or a waning rate, subsequent to the originating stimulus. The mathematical expressions for these pre-conditions were the differential equations for the linear logistic for steady acting and the harmonic logistic for waning acting. Variant forms of these curves were developed. Two exploratory experiments, or pretests, comprised launching a coffee slogan in a town and imitating a badge wearer in a boys' camp. Since the activity rate waned harmonically in both cases, the harmonic logistic fit best in both the town and the camp as expected by the hypothesis.

I. *The Need for Waning Interaction Models*†

The Washington Public Opinion Laboratory is studying the principles of human interaction in the form of diffusing messages from person to person, using questions such as: How fast will a message spread (under specified conditions)? How far? How fully and how faultlessly? How effectively in arousing belief, retelling, and compliance? What conditions will maximize social diffusion? The principles should be stated in operational rules such as mathematical curves or other models. Each model should specify (a) the variables, (b) the social preconditions expressed as mathematical assumptions, (c) the consequent curve or formula which expresses their expected joint functioning, and (d) the procedures for testing the fit of the model. All this is a case of applying dimensional methods of analysis in the field of social physics (1, 2).

This paper deals only with the questions: How fast will a message

*A paper read before Section K sponsored by the Committee on Social Physics of the AAAS Conference in Boston, December 30, 1953. This research was supported in part by the United States Air Force under Contract AF 33(038)-27522, monitored by the Human Resources Research Institute (now, Officer Education Research Laboratory, Air Force Personnel and Training Research Center), Air Research and Development Command, Maxwell Air Force Base, Alabama. Permission is granted for reproduction, translation, publication and distribution in part and in whole by or for the United States Government.

†For the Air Force, the project seeks to improve the leaflet weapon in psychological warfare. The Air Force needs to know how to maximize the desired effects of the leaflets they will drop (and have dropped by the millions) on enemy or captive populations or on our own population in an emergency. For a published description of this project and some of its findings, see (4–9); for the interactance hypothesis and dimensional theory, a special case of which is presented in this paper, see (1–3).

spread? What will be its growth curve relating people-knowing-the-message to time elapsed when all other factors are constant? It is limited to two cases where an all-or-none act spreads through a homogeneous population whose acting is either (a) steady in time or (b) waning since the stimulus that started it. Pretests indicate that the activity of retelling the message tends to wane inversely with time. Since these observations seem to call for either linear logistic curves or harmonically waning logistic curves, this paper is a report on developing and testing such models.

II. *The Social Preconditions*

The social preconditions that are assumed in the simplest case may be stated in terms of three factors (at their first powers): (1) a human *population*, P; (2) an *activity*, A; (3) a period of *time*, T.

The relation at issue is that of the population-showing-the-activity to time (i.e., $PA^0 \alpha T$). How does the number of knowers of the message grow with the time elapsed since the message started?

The preconditions should also be stated in terms of these three factors only. This proves possible by stating them in terms of maximizing a proportion of the *zero*-th statistical moment and minimizing second moments or powers of these factors. The preconditions assumed may be loosely stated here (and operationally further defined with testable indices below) as follows:

1. The population interacts randomly. Everyone has an equal opportunity to interact—a meeting being determined by *many*, *small*, and *uncorrelated* influences. It seems likely that this condition may be approximated in a population that is sufficiently large and homogeneous for probability principles to work out smoothly.
2. The time is sufficiently long to observe most of the growth or diffusing of the activity, i.e., from two hours up to three days.
3. The activity is a novel all-or-none act of any person upon any other person. The first occurrence of a one-way all-or-none activity is chosen at first for simplicity in studying interaction, here the retelling of a message.
4. The activity rate or "potency rate," defined as acts-per-actor-per-period, is either: (a) steady from period to period—the linear case; or (b) waning with the time elapsed since the start—the harmonic case.

The acting may be expected to be steady when the stimulation is steady for everyone, as in the case of some dated goals ahead. It may be expected to wane with time under one or more of at least three possible causal mechanisms which might be labeled (a) *overlaying*, (b) *dropping out*, and (c) *rebuffing*.

Whenever an event is the unrepeated stimulus and it becomes overlaid with other somewhat equivalent events or interests in each of the n succeeding time periods, the first event will then tend to be reduced by a factor of $1/n$ in the public's attention. The public's responding to such a punctiform stimulus will then decrease inversely with the time since it happened.

The acting may also be expected to wane with time as a *drop-out* effect whenever there are individual differences in the speed and/or output of activity. If the output is fairly constant, faster actors will finish and drop out, leaving fewer and slower actors in the latter periods. Similarly, if output varies but speed is fairly constant from person to person, then those of low output will finish early and drop out, leaving progessively fewer actors. If both output and speed vary, the dropping out is accelerated and the average activity rate must wane as time goes on, the form of this curve of waning activity depending on the distribution of the individual differences.

The interacting may wane with time due to the tellers getting more and more rebuffs. As more and more people hear the message, they will forestall its teller. He will stop trying to tell the message if he gets rebuffed often enough. This explanation, like that above, will result in a slackened pace of telling and of individuals ceasing to tell or *dropping out*. But the cause in rebuffing is social, whereas the cause in the other case is more specific to the individual.

The two major conditions of human interaction here (namely, that the population meet randomly and that the activity rate be specified), are highly general. They transcend any local culture or transient situation and apply to any all-or-none novel behavior. Personal, situational, and cultural conditions may affect the numerical size of the activity rate of a particular act (or message) in a particular population and a particular situation. But if the given activity rate be either steady or waning in a large homogeneous population, then a logistic curve of growth is hypothesized to be the necessary consequence.

III. *The Mathematical Derivation of Models Matching the Preconditions*

A. The Linear Logistic. The derivation of the linear logistic curve assumes a large population (at least over 100 and preferably over 1,000) which is divided into two proportions: p, the proportion of knowers, and q, the proportion of nonknowers, at any moment, so that $p + q = 1$. Let them mix thoroughly during a unit period. *This social interacting is mathematically represented by* an *overtelling* proportion, p^2, of knowers meeting knowers; a *first-telling* proportion, pq, of knowers meeting nonknowers, and a *nontelling* proportion, q^2, of nonknowers meeting nonknowers. Assuming independent probabilities, pq is the probability of a meeting in which the message can be spread. Let k represent the conditional probability of an actual telling. Thus, pq is the probability of a knower and nonknower meeting, and k is the probability of telling-if-met. k is observable from the "activity rate" or "potency rate" defined as the hearers-per-teller-per-period. The product, kpq, is then the net probability of a first telling during a unit period. This is the expected rate of growth of the message. Written as the differential

equation for an increment of diffusion in an increment of time,

$$\frac{dp}{dt} = kpq. \tag{1}$$

k is assumed to be constant from period to period as p increases and q decreases. Integrating (1) over time gives the linear logistic, a symmetric S-shaped growth curve,

$$p_t = \frac{1}{1 + \frac{q_0}{p_0} e^{-kt}} \quad \text{or} \quad \frac{p_t}{q_t} = \frac{p_0}{q_0} e^{kt}, \tag{2}$$

where p_0 and q_0 denote the knowers and nonknowers at the start, and where $k/4$ is the slope at mid-date and mid-diffusion where the slope is maximal. Thus k shows the general steepness of the curve or speed of diffusion in a general way.

This simple logistic may be generalized in many ways, one of which is to substitute a function of time, $f(t)$, for the constant k and rewrite (1) as

$$\frac{dp}{dt} = f(t)pq. \tag{3}$$

This function is k_0/t in the harmonic logistic equation below. The cumulative "augmented" logistic growth curve then is

$$p = \frac{1}{1 + \frac{q_0}{p_0} \exp\left[-\int_0^t f(x)\, dx\right]}. \tag{4}$$

A quadratic exponent giving a cubic logistic fits some of our data better than a linear logistic, but it requires four parameters instead of two, and parameters with no social interpretation at present. Still further generalizing, p and q may be replaced by integratable functions

$$\frac{dp}{dt} = f_1(t)f_2(p)f_3(q). \tag{5}$$

The subscripts here denote different functions. The dimensional family of these functions uses integral exponents, positive or negative or zero, to specify particular functions which describe many important social situations.

The linear logistic may be written in discrete form as

$$p_{t+1} = p_t + kp_tq_t\ , \tag{6}$$

where p_t is the cumulated proportion of knowers at time t, q_t the nonknowers, and $t + 1$ the next unit period.

This may be rewritten in terms of each successive proportion's being equal to the mean plus the (weighted) variance of the attribute in the pre-

ceding period, i.e.,

$$M_{t+1} = M_t + kV_t \,. \tag{7}$$

For a quick trial-fitting by plotting the data on semi-logarithmic paper, the rectified logistic is convenient:

$$\ln p/q = kt + \ln p_0/q_0 \,. \tag{8}$$

If such a plot is linear, the slope, k, is the "potency" parameter.

The logistic curve is factorable, since it is a simple product of the waxing exponential growth curve,

$$p_t = p_0 e^{k_1 t}, \tag{9}$$

times the waning exponential growth curve,

$$1/q_t = 1/q_0 e^{k_2 t} \qquad (\text{or } q_t = q_0 e^{-k_2 t}). \tag{10}$$

The right-hand side of equation (3) is their product when $k = k_1 + k_2$ (see Figure 1).

The harmonic logistic is similarly factorable into the two harmonic exponential curves. All of these are factorable in both their differential equation and their integrated forms.

A special variant form of the logistic becomes the Gompertz or "simplex" growth curve. The cumulated discrete linear logistic when $k = 1$ is also a special case of the Gompertz curve with a growth rate of 2. Thus (6) can be rearranged as

$$1 - q_{t+1} = p_t + p_t - p_t^2$$

or

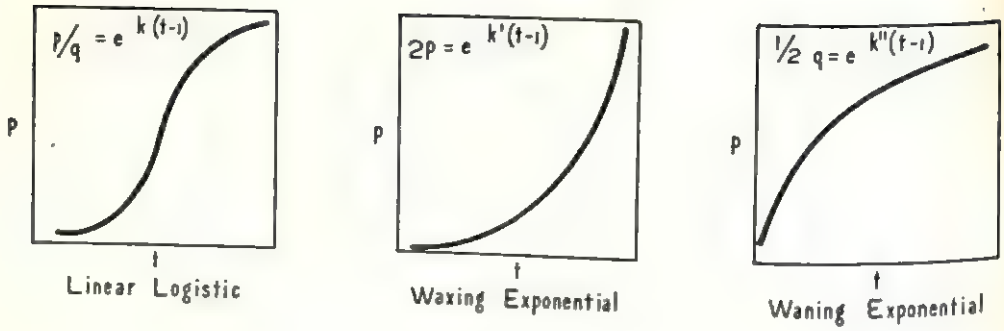
$$q_{t+1} = 1 - 2p_t + p_t^2 = (1 - p_t)^2 = q_t^2 \,.$$

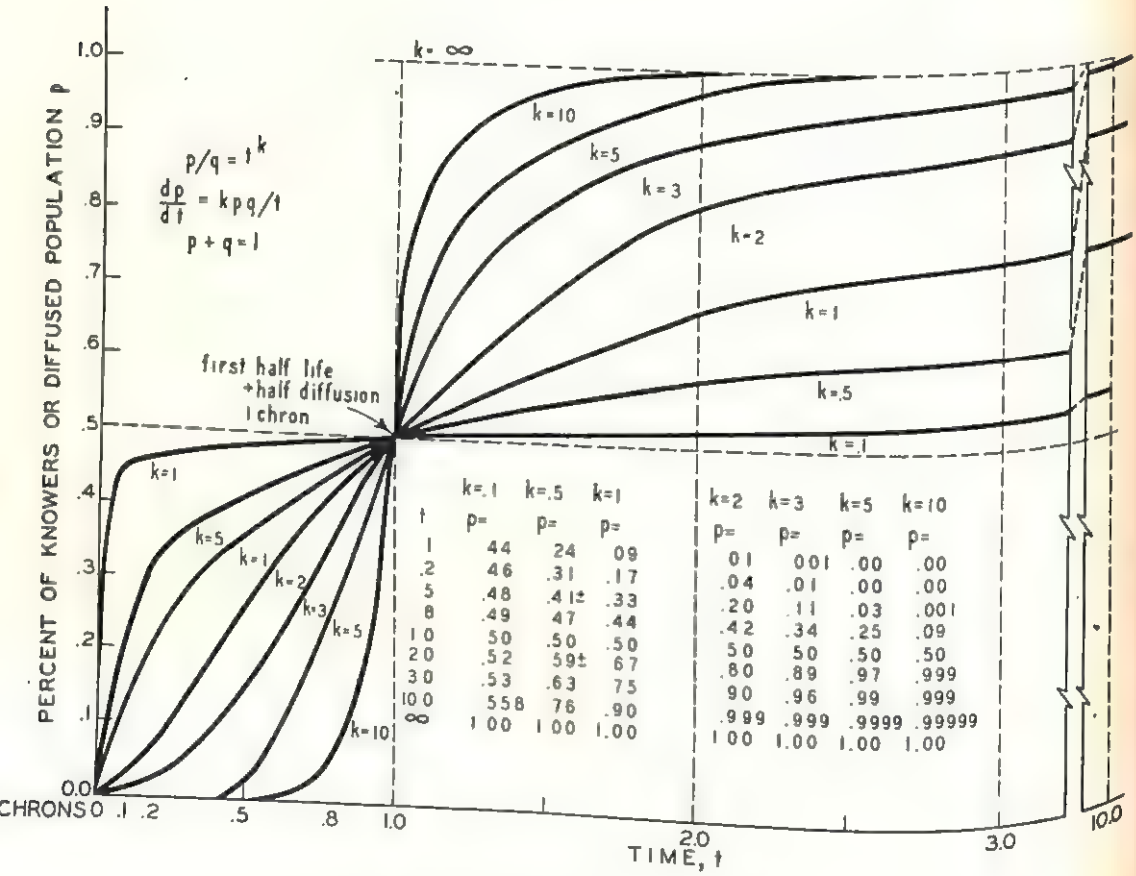
Starting with q_0 and squaring in each of the t successive periods gives

$$q_t = q_0^{2^t} \,. \tag{11}$$

This describes the decreasing of nonknowers, since q is less than unity and the exponent is greater than unity. If there are 99 per cent nonknowers at the start (q_0), it takes some three unit periods to shrink them to less than one per cent, i.e., $(.99)^{2^{3+}} < .01$.

B. The Harmonic Logistic. Next, if the activity rate k is not constant but is observed to reduce toward zero in time, the simplest form of descriptive curve is a harmonic series or inverse integers from one on up (or a hyperbola for continuous data). A straight line is simpler but cuts the axis and becomes meaningless when negative. The harmonic curve is preferable, since it fades away asymptotically. Its cumulative form is in terms of natural logarithms. It has many social applications. One such is Zipf's size-rank rule and the





t	k=.1 p=	k=.5 p=	k=1 p=
1	44	24	09
.2	46	.31	.17
5	.48	.41±	.33
8	.49	47	.44
1 0	50	.50	.50
2 0	.52	59±	67
3 0	.53	.63	75
10 0	558	76	.90
∞	1 00	1 00	1.00

k=2 p=	k=3 p=	k=5 p=	k=10 p=
01	001	.00	.00
.04	.01	.00	.00
.20	.11	.03	.001
.42	.34	.25	.09
50	50	.50	.50
.80	.89	.97	.999
90	.96	.99	.999
.999	.999	.9999	.99999
1 00	1.00	1.00	1.00

The Harmonic Logistic Growth Curves in Chron Units

FIGURE 1

hypothesized least effort principle. In cumulative form, it can often be interpreted as an extension to new fields of the Weber-Fechner relation. Such theoretical considerations, together with the fact that a harmonic curve has fitted the time series of activity rates closely in several pretests, determined its exploration here.

The harmonically waning activity rate in the logistic equation for interacting is specified by

$$\frac{dp}{dt} = k_0pq/(t + 1), \tag{12}$$

where k_0 is the potency rate at time 0 and $k_0/(t + 1)$ is the potency rate at any time t. The $t + 1$ means choosing $+ 1$ as the time origin to avoid having the growth rate become infinite at $t_0 = 0$, as it would in the simpler form, $dp/dt = kpq/t$. (With the simpler form, one can use the convention that growth starts at $t = 1$, not at $t = 0$, to avoid an infinite growth rate.) This integrates to the cumulative growth curve,

$$p_t = \frac{1}{1 + \frac{q_0}{p_0}(t + 1)^{-k_0}} \quad \text{or} \quad \frac{p_t}{q_t} = \frac{p_0}{q_0}(t + 1)^{k_0}. \tag{13}$$

Here p_t denotes the cumulative proportion of knowers at time t and the zero subscript denotes the starting moment (see Figure 1).

C. Units, Range, and Inflection Points. In order to standardize the abscissa units of time in the asymmetric harmonic logistic curve and so make different growth curves more comparable, the first "half life" may be taken as a standard unit. This "chron," as it may be called, is the time from *no* diffusion to *half* diffusion, from zero per cent of knowers, p, to 50 per cent of knowers. (The quarter life or any other fractional life might be used instead but with loss of simplicity.) The first "half life" does not equal the second. Since this curve is asymptotic at the upper limit only, the chron, or the period for the first 50 per cent to become knowers, is definite, while the remaining period is indefinite. In terms of chrons (with the origin at the absolute zero of diffusion) (13) becomes

$$\frac{p}{q} = t^{k_0}. \tag{14}$$

The chron may be viewed as an inverse function of the potency of the message (when comparing populations of the same size), for it is a reciprocal of an activity rate. The activity rate states "acts-per-period;" the chron states a "period-per-half-the-acts," or a time for 50 per cent of the population to be diffused. The longer the chron, therefore, the more "impotent" is the message. Thus, the chron or first half life becomes another standardized measure of strength of a message in a given population and situation. But

it is not a simple inverse potency, for it is stated in terms of a *per cent* of a total population. Thus, chrons as a measure of message strength are comparable in different populations only insofar as the populations are alike in size. Different potency rates, however, are comparable from any population since they are per capita activity rates.

The reasons for using the chron unit are simplicity of the formula and flexibility of fitting. In chron units the formula drops the explicit p_0/q_0 term and can be written in simplest form as $p/q = t^k$, or as $p = qt^k$. In chron units the node where all the family of curves intersect (Figure 1) is at the half diffusion point where $p = q = .5$. This gives a wide range of shapes from convex-up to S-shape available to fit given data. In smaller time units the node occurs so far to the left, or so early in the total growth period, that the curves are convex-up for most of their range. In that case, a harmonic logistic model might *apparently* give a bad fit even though the growth was logistic and its activity rate, k, did wane harmonically. This misleading result can happen when the time continuum is subdivided into a size of time unit and measured from an origin point which are inappropriate and obscure the harmonic waning of the potency or activity rate. The chron is an "optimal" time unit in the sense that its length is so fitted to the data in hand as to develop a versatile family of curves among which a close fit to the data may be more attainable.

In practice, the range between the upper and lower limits, or asymptotes, varies with the potency rate. The time range in the linear logistic is from $+\infty$ to $-\infty$. In practice, it has to be truncated at arbitrary points (such as p = one per cent and p = 99 per cent), since the curve approaches 100 per cent and zero per cent asymptotically as time goes to plus or minus infinity. The harmonic logistic, however, starts from absolute zero point in time (with no knowers, so $p_0 = 0$ at t_0) and goes to 100 per cent asymptotically. For some purposes it may be truncated at 10 chrons, giving a standard time range of 100 decichrons along the abscissa to match the population range of 100 percentage points along the ordinate. Figure 1 shows these values of p at 10 chrons for various potencies.

The zero point of growth in the harmonic logistic requires extrapolating to determine it accurately and in order to fit data more closely. For the starters' date, t_A, is after the absolute zero, t_0, from which the chron should be computed. The starters come into the curve, as it were, at the date in chron units when it has grown up to the starters' percentage of knowers, p_A (on the curve fixed by one k).

A difficulty in fitting these qt curves is to determine the absolute zero point, t_0, in time, which occurs when p, the growth, is zero, p_0. But in practice the first observable amount of growth is when the starters, p_A, become knowers, which is later than the true zero point by a small but unknown amount. To estimate the zero date, the starters' conversion date

may be taken as a first approximation for computing the chron and then k from the slope of the points. Using this best-fit estimate of k and the starters' proportion, p_A, as p, one can then solve for t, getting a second approximation. Further approximations may be made by repeating these steps until the correction yielded is negligible. It seems likely that if the starters are few (such as less than one per cent), the error in the chron will usually be of the order of one per cent for values of k near unity. Then a single correction will be sufficient. It will always put the starters' date at some positive fraction of a chron after the estimated absolute zero date and thus will lengthen the chron and flatten the growth curve slightly.

The upper limit of the growth curve always presents a problem. Whether chrons or other time units are used, in the case of either the linear or the harmonic logistic, it is necessary to know P_∞, the terminal population, in absolute numbers. The difficulty is that there may be many different interpretations of P, such as:

1. The *census population* in a diffusion area (which may be too large, since it may include undiffusable elements such as babies who cannot talk).
2. The *relevant population* in a diffusion area which is thought a priori to be diffusable but which may have an undiffusable fraction cut off from each other by unknown barriers of physical, physiological, psychological, or cultural origin.
3. The final *diffused population*, which may be unknown and have to be estimated as generally less than the relevant population. If the latter (such as "all adults") is known, it logically should yield better predictions than the larger census population. But insofar as the diffused population differs from the researcher's a priori judgment of the relevant population, the fits will be loose and the predictions poor.

We have discussed two harmonic logistic curves, namely,

$$\frac{dp}{dt} = kpq/(t + 1), \tag{15}$$

where t is expressed in some conventional time unit, and

$$\frac{dp}{dt} = kpq/t, \tag{16}$$

where t is expressed in chron or half life units. In the case of (15), we find that the shape of the curve has the following dependence upon k: For $p_0 < \frac{1}{2}$, the curve is concave downward throughout; for

$$0 < k \leq \frac{1}{1 - 2p_0}$$

and for

$$k > \frac{1}{1 - 2p_0},$$

the curve is sigmoid with an inflection point at

$$t' = \left(\frac{q_0}{p_0} \cdot \frac{k-1}{k+1}\right)^{1/k} - 1. \tag{17}$$

In the case of (16), where time is expressed in chron units, the curve is concave downward throughout for $k \leq 1$ and is sigmoid for $k > 1$, with a point of inflection at

$$t' = \left(\frac{k-1}{k+1}\right)^{1/k}. \tag{18}$$

The harmonic logistic family of curves thus has a wide range of shapes, determined by k, combined with any steepness as determined by the chron, c.

D. "Timeless" Forms of the Harmonic Equations. It proves possible to rewrite the harmonic equations above explicitly in terms of k, p, and q alone with the time factor "cancelled out." This means that the growth can be described and predicted in terms of the knowing and nonknowing proportions exclusively without knowing whether the time units are ordinal units of removes from the first teller or cardinal units of clock time. For this, substitute the t in the equations above into their respective differential equations:

Curves	*In general*		*For $k = 1$*	
Harmonic logistic	$\frac{dp}{dt} = k_0 p^{1-(1/k_0)} \cdot q^{1+(1/k_0)}$	(19)	$\frac{dp}{dt} = q^2$	(19a)
Harmonic waxing exponential	$\frac{dp}{dt} = k_1 p^{1-(1/k_1)} \cdot \left(\frac{1}{2}\right)^{1/k_1}$	(20)	$\frac{dp}{dt} = \frac{1}{2}$	(20a)
Harmonic waning exponential	$\frac{dp}{dt} = k_2 q^{1+(1/k_2)} \cdot \left(\frac{1}{2}\right)^{-1/k_2}$	(21)	$\frac{dp}{dt} = 2q^2$	(21a)

Note that the sum of exponents on the two population factors in (19) is 2 for the logistic. Dimensionally this denotes *interaction*, i.e., a group phenomenon. The sum of the exponents in (20) and (21) is one. Dimensionally this denotes *action* (or reaction), i.e., a plurel phenomenon.

The empirical or social interpretation of (19a), (20a), and (21a) in the special case when $k = 1$ has not yet been fully explored. It appears that (20a) reflects constant growth, because the curvature of the waxing

exponential is here exactly counterbalanced by the opposite curvature of the harmonic activity rate. In (21a), the harmonically waning exponential curiously has just twice the growth rate of the harmonic logistic in (19a). One possible social interpretation of the logistic (19a) having q^2 as its growth rate seems at present to be that the growth rate depends on two factors, one physical and one psychological, which here both happen to be numerically the same, namely, q. Thus, the growth rate shrinks as q shrinks when nonknowers become fewer and physically harder to find. But, in addition, a psychological factor may operate in that as the tellers get more and more rebuffs from the growing proportion of knowers, the tellers slacken their telling activity and this slackening keeps pace with q, the proportion of current nonknowers. This explanation and alternative ones must be empirically tested.

IV. *Some Experimental Testing*

In order to begin testing the foregoing theory and to develop more definite tests, a dozen preliminary experiments were made in the first year of Project Revere. Two of these will be reviewed here as tests of the harmonic logistic hypothesis. [Some of our tests of the linear logistic have been published elsewhere (**7**, **8**, **9**)]. Both were designed to test linear logistic models in clock time units. In the first, however, the activity rate, k, was found a posteriori to wane harmonically with removes, and its growth from remove to remove should therefore fit the harmonic logistic more closely. In the second set of data to be reported, the activity rates were also observed to wane, and therefore the harmonic logistic should again give a posteriori a better fit than the linear logistic.

Time measured in removes or generations of hearers is in ordinal units and is more free of diurnal and other rhythms which "overlay" the growth curves in clock or cardinal time units. Ordinal units seem apt to yield smoother curves which fit models more readily. But curves in cardinal time units are needed wherever practical prediction of growth in clock time is wanted. In both cases the reward offered as stimulus was expected to evoke steady acting throughout the whole period. Instead it produced a spurt of activity which waned steadily—perhaps because of the overlaying or drop-out, or rebuffing mechanisms noted above.

A. Coffee Slogan Diffusing in a Town. One randomly chosen housewife of every six in a village of 950 inhabitants was told a new coffee slogan by an interviewer ringing her doorbell on Monday morning. All were invited to retell it to their friends. A free pound of coffee was promised for every housewife in town who might know the slogan when every household would be canvassed later.

This message spread from person to person till 88 per cent of the housewives knew it on Wednesday's census of households, determining the pro-

portion of knowers, the ordinates of the curve, or P values in Table 1, Column 3. In this census questions of who told whom, when, and where identified the remove of each respondent and so measured the increment of new knowers at each remove of retelling and the potency of hearers-per-teller of each remove. The potency was found to wane with successive removes in an harmonic curve. Thus, the harmonic subcase of the logistic should fit these data better than the linear case.

TABLE 1

Data for Testing the Harmonic Logistic Growth Model in "C-ville"

Removes (ordinal time units)	Chron time	Observed cumulated population of message knowers	Observed increment of knowers	Expected increment of knowers by the harmonic logistic model
t	t_c	P	ΔP	$\Delta P'$
0	.84708	42	22.83%	21.46%
1	1.05876	111	37.50	39.52
2	1.27044	164	28.80	25.68
3	1.48212	178	7.61	8.93
4	1.69380	180	1.09	2.81
5	1.90548	184	2.17	0.96
Totals	1.9 "first half-life" units	184 housewives	100.00%	99.36%

1 remove = .21168 chrons

The observed potency rates for the successive removes showed a closeness of fit correlation coefficient of .99 with the best fitting simple harmonic curve ($a = k/t$). (See Table 1).

The correlation coefficient of the increments of the observed growth of message knowers with the increments in growth expected by the harmonic logistic curve ($p/q = t^k$) was also .99 (i.e., $r_{\Delta P \Delta P'} = .99$). By the z test this r is significantly different from zero and also from our arbitrary standard of close fit, namely, $r = .9$ at the 5 per cent level. (Exactly how applicable the z test is here, however, is unknown since the variate p, "knowing the message," was dichotomously observed and may not be normally distributed; also while the starters were a 20 per cent random sample of the households, only one town was studied, and without replication this may not be representative of other communities.)

The standard for nonrejection of a hypothesis was that (a) the closeness of fit correlation index should exceed .9 between the observed and the model-

expected data in uncumulated form, and (b) this r should be significantly different from zero at the five per cent confidence level. (This generally entails an r in cumulated data, such as is usually reported, above .99, but this test is insensitive and partly spurious since cumulating compels some correlation even in random series.)

The closeness of fit correlation of the uncumulated data to the linear logistic curve which is based on a steady activity rate was .37. This linear logistic hypothesis then was rejected. But the similar closeness of fit correlation (by a successive approximations technique in fitting) of the uncumulated data to the harmonic logistic, based on a waning activity rate, was .99. Therefore the harmonic logistic hypothesis could not be rejected.

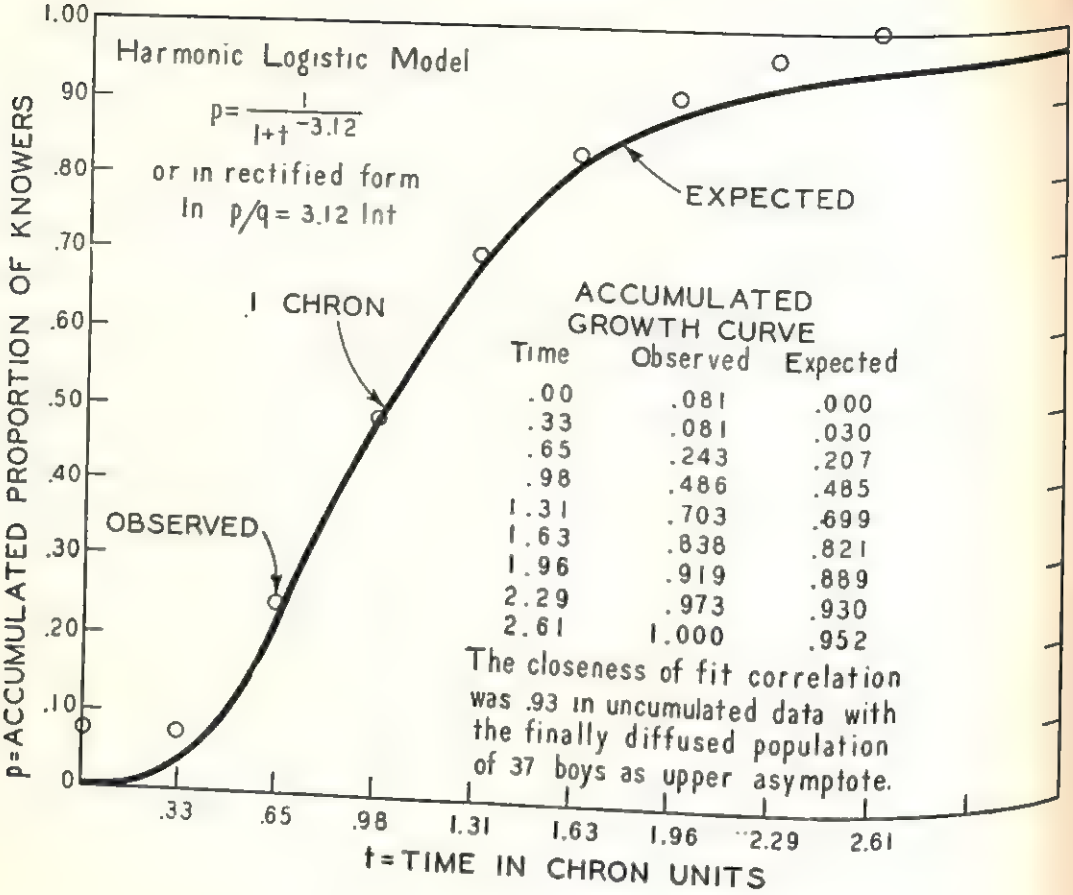
An excellent fit of one model to given data does not preclude fits as good or better by other models. In fact, the waning random net variant of the logistic, which has a constant "potency ever," fitted these uncumulated data with $r = .996$ as reported by Professor Anatol Rapoport, who developed this random net model. Also, the logistic in discrete form (6) often seems to fit our data better than the continuous curve.

B. Contagious Behavior in a Boys' Camp. Three boys in a summer camp of 42 boys were given large yellow buttons with the name of the camp and a question mark in black lettering on them. These "starters" circulated during the noon rest period and, when asked, said that they had been told a few more such buttons were obtainable at the lodge where they received their buttons. The exact time at which each boy came in to ask for a similar button was recorded and so the growth curve could be plotted. The growth rate of hearers-per-teller waned in the successive 15-minute periods during the two hours in which 39 of the boys came in for buttons. The total growth data were fitted to both linear and harmonic logistic curves (Figure 2). The closeness of fit correlations were computed on uncumulated data and were significantly different from zero at the one per cent confidence level. The correlation was .88 for the linear logistic and .93 for the harmonic logistic. The slightly better fit of the harmonic is in line with the theoretical expectations but the difference was too small and the underlying population also too small to warrant much dependence upon these findings.

For both of the tests described here, the chi square test showed that the discrepancies between model and data were not significant at the five per cent level:

	First Pretest	*Second Pretest*
Chi square	6.163	1.660
Degrees of freedom	4	6
Probability	$.90 < p < .95$	$.10 < p < .20$

We conclude that in both pretests the looseness of fit, i.e., discrepancy of model and data, was both descriptively small and statistically not significant



Diffusion of a Message in a Boys Camp Population
Fitted by a Harmonic Logistic Growth Curve in Chron Units

FIGURE 2

at the five per cent level. The discrepancies here are unimportant practically and may be due to sampling error.

The tests above were concerned with "hypotheses of form," not "hypotheses of amount." These hypotheses asserted the *form* of relation between the variables as being a linear logistic curve, etc., and did not assert the *amount* of each parameter of that curve—which could not be expected in wholly new situations.

The amounts or sizes of the parameters were determined by least squares techniques to find the best fit. Then the Pearson correlation coefficient was used to measure how closely this best fitted curve corresponded to the observed data. The technique of fitting matches the mean of the model to the mean of the data and similarly matches the two variances, leaving only

the variable discrepancy to be measured by the correlation. For this reason, the Pearson r is here almost identical with the intraclass r. The latter r is the more exacting descriptive statistic of closeness of fit, since it can approach unity only if the mean, variance, and rank order of the data agree with these moments in the model.

REFERENCES

1. Dodd, Stuart C. Dimensions of society. New York: Macmillan, 1942.
2. Dodd, Stuart C. Systematic social science. American University of Beirut, Social Science Series No. 6, 1947. (University Bookstore, Seattle, agents).
3. Dodd, Stuart C. The interactance hypothesis: a gravity model fitting physical masses and human groups. *Amer. sociol. Rev.*, 1950, **15**, 245–256.
4. Dodd, Stuart C. On all-or-none elements and mathematical models for sociologists, *Amer. sociol. Rev.*, 1952, **17**, 167–177.
5. Dodd, Stuart C. Testing message diffusion from person to person. *Public Opinion Quarterly*, 1952, **16**, 247–262.
6. Dodd, Stuart C. Testing message diffusion in C-ville. Proceedings of the Pacific Sociological Society, *Research Studies of the State College of Washington*, 1952, **20**, 83–91.
7. Dodd, Stuart C. Testing message diffusion in controlled experiments. *Amer. sociol. Rev.*, 1953, **18**, 410–416.
8. Dodd, Stuart C., and Winthrop, Henry. A dimensional theory of diffusion—an analysis, modeling and testing of one-way interaction. *Sociometry*, 1953, **16**, 160–202.
9. Dodd, Stuart C. Can the social scientist serve two masters—an answer through experimental sociology. Proceedings of the Pacific Sociological Society, *Research Studies of the State College of Washington*, 1953, **21**, 195–213.
10. Dodd, Stuart C. Diffusion is predictable. *Amer. sociol. Rev.* In press.

Manuscript received 3/26/53

Revised manuscript received 5/9/55

PSYCHOMETRIKA—VOL. 21, NO. 2
JUNE, 1956

A METHOD FOR OBTAINING AN ORDERED METRIC SCALE*

SIDNEY SIEGEL

PENNSYLVANIA STATE UNIVERSITY

A method is presented for collecting data which will yield a scale on which the entities are ranked in preference (ordinality), the distances between the entities on the scale are ranked (ordered metric), and all combinations of the distances are ranked (higher-ordered metric). The sources drawn upon are von Neumann and Morgenstern (9), and lattice theory. An empirical example is given in which a higher-ordered metric scale is derived.

If an individual is both consistent and transitive in his preferences with respect to a group of discriminable entities, it is possible by the method of paired comparisons to rank these entities in the sense of an ordinal scale of utility, e.g., $A > B > C > \cdots > N$ (read: A is preferred to B, etc.). An individual is *consistent* if he prefers the same entity of a pair whenever that pairwise comparison is presented to him. His preferences are *transitive* if when $A > B$ and $B > C$, then $A > C$. The entities involved (A, B, C, etc.) may be objects or actions. The *utility* of an entity is, roughly, the subjective value of that entity.

Such a scale, however, gives no information about the relative sizes of the differences in utility between the entities. Coombs (**2**, **3**, **4**), among others, has shown that knowledge of the magnitudes of these differences would strengthen the measurement of the psychological attribute involved and, therefore, increase the amount of information obtained from the responses made by the individual to the stimuli presented to him.

Coombs (**2**) suggests the label *ordered metric* for those scales which give not only an ordering of entities but also at least a partial ordering of the distances between the various entities. Coombs also presents a method, which he calls the *unfolding technique*, for obtaining an ordered metric scale (J scale) from a rank order preference scale (ordinal I scale).

A method is developed in this paper for obtaining an ordered metric scale of preference. This method is particularly suitable for the measurement of utility, which is a central concept in decision theory. The ordered metric

*I am grateful to Professor William L. Lepley (Department of Psychology) and Professor Jack R. Tessman (Department of Physics) for their critical reading of this paper. Paul Hurst and Robert Radlow participated in many discussions on the form of measurement discussed in this paper, and assisted in collecting data. I am also grateful to Professor T. C. Benton (Department of Mathematics) for certain source materials.

scale derived by this method yields the following information:

a. It orders the entities involved, i.e., $A > B > C > D > E \cdots$.
b. It orders the distances between the entities, i.e., say $\overline{DE} > \overline{AB} > \overline{CD} \cdots$, where $\overline{AB}$ implies the difference in utility between A and B.
c. It orders all possible combinations of contiguous distances between entities, i.e., say $\overline{AB} + \overline{BC} > \overline{BC} + \overline{CD} + \overline{DE}$, or $\overline{AB} > \overline{CD} + \overline{DE}$.

This method is not restricted by the number of entities to be scaled.

Because this method yields more than a partial ordering of the distances between the entities (see *b* and *c* above), it is suggested that this type of scaling be termed *higher-ordered metric*. Coombs (**4**) seems to mean this type of scaling by his term *ordered ordered*. However, a scale may be *ordered ordered* and still not satisfy *c* above. (The author has resisted the temptation to call the present technique *ordered ordered ordered*.)

The author has collected data which permit a higher-ordered metric scaling of preferences for up to seven entities; however, this paper will present only those data concerned with preferences among five entities. The entities employed were books, phonograph records, or money.

Sources of the Method

Von Neumann and Morgenstern (**9**, p. 17) have suggested that measurement of a person's utility in a stronger sense than ordinality could be obtained if

a. the person can always say whether he prefers one entity to another, and
b. the person can also completely order probability-combinations of the entities, i.e., combinations of entities with stated probabilities of attainment, e.g., $(B, A; p)$—read: the combination of B and A, with probability p of getting B, and probability $(1 - p)$ of getting A.

Condition *b* requires some explanation. Suppose for a given individual $A > B > C$. This individual is given a choice between $(A, A; 1/2)$, i.e., getting A for sure, and $(B, C; 1/2)$, i.e., getting B if say head occurs on the toss of a fair coin, and getting C if tail occurs. It is clear that the individual will prefer the first alternative, which is $(A, A; 1/2)$, since $A > B > C$. In probability-combinations of entities, say $(x, y; 1/2)$, the prospect is of getting *either* x or y. The probability of getting x is .50; the probability of getting y is the remaining .50. The two alternatives in the probability-combination are mutually exclusive; the individual is absolutely certain of getting either x or y if he chooses that probability-combination.

We expect the individual to possess a clear understanding of his preferences among the entities A, B, and C (this is condition *a*), and we also expect him to prefer getting A for sure to a 50-50 combination of B and C.

Now suppose that the individual must choose between $(B, B; 1/2)$ and

$(A, C; 1/2)$. That is, he must choose between getting B for sure or getting a 50-50 chance at A or C. By making this choice, he yields new information. If he chooses the combination which gives him B for sure, his choice indicates that B is closer to A than it is to C. If he chooses $(A, C; 1/2)$, then B must be closer to C than it is to A. This is fundamentally new information because the statement $A > B > C$ told us nothing about the distances (differences) between the entities on the utility scale. Thus, the von Neumann and Morgenstern suggestions imply the possibility of measurement of utility on at least an ordered metric scale.

The second source drawn upon in higher-ordered metric measurement is lattice theory. This source is not so centrally important as the first, but it does offer a heuristic device for indicating the minimum information necessary for achieving higher-ordered metric scaling. Birkhoff (**1**, p. 6, pp. 66-72) suggests various diagrams which give a descriptive ordering of entities. Coombs (**4**, p. 4; **5**, p. 475) suggests such diagrams and puts them to use.

The lattice used here (Figure 1) not only gives a descriptive ordering of probability-combinations of entities (based on the individual's preference rankings) but also makes apparent which probability-combinations are not orderable from just a knowledge of the preference rankings. Such probability-combinations will be called *non-orderable.*

If an individual's preferences among five entities are $A > B > C > D > E$, the probability lattice is shown in Figure 1; where there is a connecting line between two probability-combinations, the higher probability-combination is preferred to the lower (**5**, p. 475). In other words, if it is true that $A > B > C > D > E$, then any two probability-combinations on the lattice that can be connected with a line which is consistently going up (or down) can be ordered, with the higher probability-combination being preferred to the lower, e.g.,

$$(A, E; \tfrac{1}{2}) > (B, E; \tfrac{1}{2})$$

$$(A, D; \tfrac{1}{2}) > (A, E; \tfrac{1}{2})$$

$$(B, D; \tfrac{1}{2}) > (D, D; \tfrac{1}{2}), \text{ etc.}$$

Simple ranking tells us nothing about the non-orderable relations, i.e., any two probability-combinations which cannot be connected by a line always going in the same direction (with respect to the horizontal-vertical dimension), e.g.,

$$(A, E; \tfrac{1}{2}) \ ? \ (B, D; \tfrac{1}{2})$$

$$(A, E; \tfrac{1}{2}) \ ? \ (B, C; \tfrac{1}{2})$$

$$(A, E; \tfrac{1}{2}) \ ? \ (B, B; \tfrac{1}{2})$$

$$(A, D; \tfrac{1}{2}) \ ? \ (C, C; \tfrac{1}{2}), \text{ etc.}$$

It is the relations between these (non-orderable) pairs of probability-combinations which contain the information necessary to change an ordinal scale to a higher-ordered metric scale.

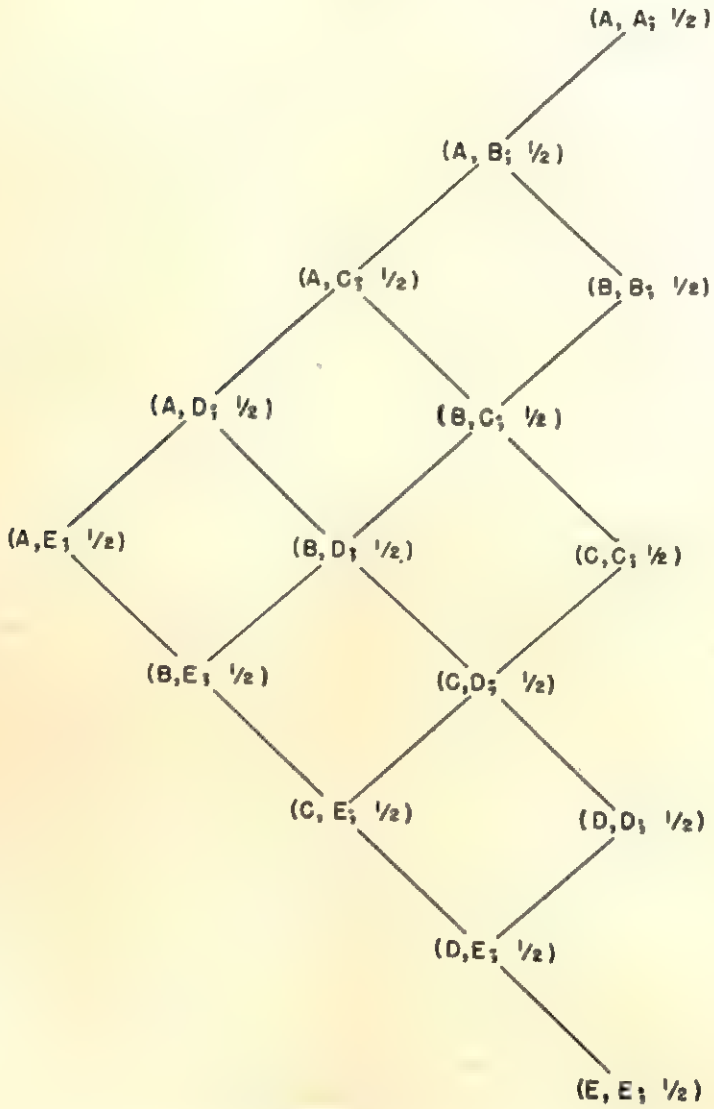


Figure 1.

A careful perusal of the lattice in Figure 1, which is based on five entities, will disclose that there are 15 pairs of probability-combinations which cannot be connected by an always-rising or always-descending line. It can be shown that if N is the number of entities to be scaled, then $\binom{N+1}{4}$ gives the number

of non-orderable pairs of probability-combinations. In the case of five entities, $\binom{6}{4} = 15$.

Not all of the non-orderable relations must be found in order to obtain the information necessary for higher-ordered metric scaling. In the present example from three to six of these relations (depending on the type of underlying ordered metric scale) are needed to achieve ordinal ordered metric scaling, and ten at most are required to achieve higher-ordered metric scaling (in which all combinations of the distances can be ordered). The fact that not all 15 pairs are needed is important. The ordering of the remaining pairs can be predicted after the minimum are used to obtain the necessary information; the success or failure of these predictions provides a check on whether higher-ordered metric scaling has in fact been achieved.

In what follows the general method of obtaining a higher-ordered metric will be outlined. Then the operational definitions necessary for implementing the general method will be given. Finally an example of the empirical use of the method will be given.

General Requirements

1. Interpret the p in the probability-combinations, e.g., $(x, y; p)$ to be *subjective* probability. [Ramsey (8) first suggested the importance of subjective probability (degree of belief) in the measurement of utility. His theory resembles that given here in many respects and antedates von Neumann and Morgenstern by more than a decade. However, the latter authors were the first to make operationally clear a method for the measurement of utility. I am indebted to Professor Donald Davidson of Stanford University for demonstrating the significance of Ramsey's work to me.]

2. Find an event for which the person's subjective probability can be experimentally determined to be one-half.

3. Require the person to rank, by the method of paired comparisons, the entities used.

4. Require the person to state his preference between each non-orderable pair of probability-combinations. (As shown above, the majority of pairs of probability-combinations may be ordered, as in the lattice in Figure 1, from a knowledge of the person's ranking of the entities. Step 4 is concerned with those pairs which cannot be ordered from this knowledge.)

5. Observe those choices which will permit the determination of an ordered metric scale.

6. Observe those choices which will permit the determination of a higher-ordered metric scale.

7. Check whether the remaining choices are consistent with the scale derived from Step 6. If all of these choices are consistent, then all of the previously non-orderable (in 4 above) pairs of probability-combinations are

consistent with each other and can be predicted from the higher-ordered metric scale derived. If Step 7 succeeds, then higher-ordered metric scaling has in fact been achieved.

Operational Definitions

Davidson, Siegel, and Suppes (6), in a study designed to measure the utility of money in the sense of an interval scale, developed an event which, for most people, has a subjective probability of one-half. Such an event is difficult to find because of the prejudices and superstitions which many people hold concerning familiar events, e.g., heads on coins, evens on dice, etc. The event used here is produced by means of specially-made dice. On three faces of the die, the nonsense syllable *ZOJ* is engraved, and on the other three faces *ZEJ* is engraved. Similar dice were made with pairs *WUH*, *XEQ* and *QUG*, *QUJ* on their faces. These syllables were selected from Glaze (7), who reports these pairs to have practically zero association value. The dice were tested with each subject; in every case the expectation of zero association was upheld, i.e., each subject was indifferent about which nonsense syllable he would bet on or which one would be the winner. The use of these dice will be discussed further in the following section.

Careful and considered choices between the probability-combinations presented to a subject were assured by "realistic" conditions. That is, when books were used as entities to be scaled, the subject was assured of getting a book or books. The identity of the book he received was a function of *all* of his choice behavior. Therefore he was highly involved in each choice. When amounts of money were used as the entities, the subject was given a sum of money (usually one dollar) at the start of the session. He gambled with that money, keeping all funds in his possession at the end of the session.

The essential device that defines operationally how the subject's choices determine ordered metric scaling is a one-person game (6) in which the subject chooses between two alternatives, each of which is a probability-combination of two outcomes. The format for each offer is:

	Alternative 1	*Alternative* 2
If event E occurs:	you get w	you get x
If not-E occurs:	you get z	you get y

The subject chooses the column; the outcome of event E determines the row. Event E might be *ZOJ*, in which case *not-E* would be *ZEJ*.

Suppose $w > x > y > z$. If the subject chooses alternative 1, then

$$(w, z; p) > (x, y; p). \tag{1}$$

If $u(w)$ is read as "the utility of w" and is interpreted as the subjective value of w, i.e., its worth to the person, then (1) can be written

$$p \cdot u(w) + (1 - p) \cdot u(z) > p \cdot u(x) + (1 - p) \cdot u(y). \tag{2}$$

If p is understood to be subjective probability, and is known to be one-half, then (2) can be written

$$u(w) + u(z) > u(x) + u(y), \tag{3}$$

and

$$u(w) - u(x) > u(y) - u(z); \tag{4}$$

i.e.,

$$\overline{wx} > \overline{yz}, \quad \text{when} \quad (w, z; p) > (x, y; p). \tag{5}$$

That is, w and x differ in utility more than y and z.

It should be noted that the distances are *directed* distances. That is, $\overline{xw}$ is the negative of $\overline{wx}$. To simplify comparisons, the convention has been adopted of always deriving the distance from the more preferred to the less preferred entity. [For example, from (3) we could get $u(z) - u(y) > u(x) - u(w)$. But since $w > x > y > z$, we multiply through by -1 to get (4) as shown.]

An Example of Higher-Ordered Metric Scaling

A graduate student in psychology served as the subject. He was shown a collection of books and was asked to choose from them the five books which he would most like to own. He was told to make this selection carefully, for he would surely receive one of the books he chose at the conclusion of the session. The books he chose, in the order of choice, were: (1) S. S. Stevens (Ed.), *Handbook of Experimental Psychology*, (2) E. G. Boring, *A History of Experimental Psychology*, (3) E. R. Hilgard, *Theories of Learning*, (4) E. R. Hilgard and D. G. Marquis, *Conditioning and Learning*, and (5) H. B. English, *A Student's Dictionary of Psychological Terms*, 4th edition.

All possible pairs of these five books were presented to the subject orally, and he was asked to state his preference as each was presented. His choices were:

Stevens > Boring ($A > B$)
Stevens > Hilgard ($A > D$)
Stevens > Hilgard and Marquis ($A > C$)
Stevens > English ($A > E$)

Boring > Hilgard ($B > D$)
Boring > Hilgard and Marquis ($B > C$)
Boring > English ($B > E$)

Hilgard and Marquis > Hilgard ($C > D$)
Hilgard and Marquis > English ($C > E$)

Hilgard > English ($D > E$)

The subject's choices were consistent and transitive; his choices would be ranked thus: $A > B > C > D > E$.

Having stated his preferences among the paired comparisons, the subject was introduced to the "game." He was allowed to become familiar with the

dice and the game by taking practice trials in which simple relations were offered, i.e., those relations which are connected by a line in the lattice [e.g., $(A, D; 1/2)$ or $(B, E; 1/2)$]. As a trial run he could choose between a 50-50 chance of getting either Stevens (A) or Hilgard (D), or a 50-50 chance of getting either Boring (B) or English (E).

The practice trials served not only to introduce the game but also to check on the consistency and transitivity of the lattice (Figure 1), because all of the choices on these simple relations should be predictable. The practice trials also served to check whether the subject's behavior was consistent with a subjective probability of one-half toward the event. This was ascertained when the subject showed indifference as to which syllable would be the winner. That is, if the *ZOJ-ZEJ* die was used and the subject was willing to make his choice between alternatives 1 and 2 (without knowing or caring which of the nonsense syllables was to be associated with which of the outcomes) it was concluded that the choice was based only on the utility of the entities involved and was independent of the particular event giving rise to the outcome.

The subject was told to consider the alternatives, announce his choice, encircle that choice on a 3×5 card, and then turn that card face down. He was told that after all sets of alternatives were presented to him, the cards would be shuffled and he would then draw one card. The alternative which he had encircled on that card would be determined by a roll of the nonsense syllable die. Thus each selection made by the subject might be the crucial one, so each one had to be made carefully.

After the practice conditions were met, critical sets of alternatives were presented to the subject. His choices are indicated by the direction of the carat. For example, the first alternative, (a), permitted the subject to choose either A or C with a 50-50 probability, or to choose getting B for sure. The direction of the carat shows that he preferred to take a 50-50 gamble on getting either A (Stevens) or C (Hilgard and Marquis) rather than to be sure of getting B (Boring). The choices were:

(a) $(A, C; \frac{1}{2}) > (B, B; \frac{1}{2})$

(b) $(B, D; \frac{1}{2}) > (C, C; \frac{1}{2})$

(c) $(D, D; \frac{1}{2}) > (C, E; \frac{1}{2})$

(d) $(C, D; \frac{1}{2}) > (B, E; \frac{1}{2})$

(e) $(B, D; \frac{1}{2}) > (A, E; \frac{1}{2})$

(f) $(D, D; \frac{1}{2}) > (A, E; \frac{1}{2})$

(g) $(A, D; \frac{1}{2}) > (B, B; \frac{1}{2})$

(h) $(A, D; \frac{1}{2}) > (B, C; \frac{1}{2})$

(i) $(B, B; \frac{1}{2}) > (A, E; \frac{1}{2})$

(j) $(C, C; \frac{1}{2}) > (B, E; \frac{1}{2})$

(k) $(B, C; \frac{1}{2}) > (A, E; \frac{1}{2})$

(l) $(A, D; \frac{1}{2}) > (C, C; \frac{1}{2})$

(m) $(D, D; \frac{1}{2}) > (B, E; \frac{1}{2})$

(n) $(C, D; \frac{1}{2}) > (A, E; \frac{1}{2})$

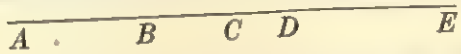
(o) $(C, C; \frac{1}{2}) > (A, E; \frac{1}{2})$

The inequalities (1) to (5) in the previous section show that these choices can be stated in terms of distances as:

(a')	$\overline{AB} > \overline{BC}$	(h')	$\overline{AB} > \overline{CD}$
(b')	$\overline{BC} > \overline{CD}$	(i')	$\overline{BE} > \overline{AB}$
(c')	$\overline{DE} > \overline{CD}$	(j')	$\overline{CE} > \overline{BC}$
(d')	$\overline{DE} > \overline{BC}$	(k')	$\overline{CE} > \overline{AB}$
(e')	$\overline{DE} > \overline{AB}$	(l')	$\overline{AC} > \overline{CD}$
(f')	$\overline{DE} > \overline{AD}$	(m')	$\overline{DE} > \overline{BD}$
(g')	$\overline{AB} > \overline{BD}$	(n')	$\overline{DE} > \overline{AC}$
		(o')	$\overline{CE} > \overline{AC}$

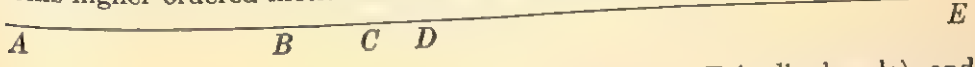
The first five relations—(a') to (e')—yield the ordered metric scale: $\overline{DE} > \overline{AB} > \overline{BC} > \overline{CD}$. When relations ($f'$) and ($g'$) are also considered, we have necessary and sufficient information for a higher-ordered metric scale. Relations (h') through (o') provide checks on the uniqueness of the higher-ordered metric scale derived by (a') through (g').

The ordered metric scale given by (a') through (e') may be depicted as:



It is seen that the subject's choices (stated in distances) in relations (h') through (l') may be predicted (i.e., checked) from a knowledge of just (a') through (e'). However, in order to predict (m'), (n'), and (o'), the subject's choices in (f') and (g') must be known; knowledge of the latter provides a more powerful form of measurement.

Inasmuch as all of the choices in the relations (h') through (o') were predictable, i.e., were consistent with choices in (a') through (g'), the higher-ordered metric scale derived from choices (a') through (g') is unique and valid. This higher-ordered metric scale may be depicted as:



We now know not only that $A > B > C > D > E$ (ordinal scale), and that $\overline{DE} > \overline{AB} > \overline{BC} > \overline{CD}$ (ordered metric scale), but also that $\overline{AE} > \overline{BE} > \overline{CE} > \overline{DE} > \overline{AD} > \overline{AC} > \overline{AB} > \overline{BD} > \overline{BC} > \overline{CD}$ (higher-ordered metric scale).

Data on a utility scale of five entities (records or books or amounts of money) have been collected from 10 subjects. Of this number, nine have been consistent; therefore it was possible to derive a higher-ordered metric scale for each of these nine. One subject showed inconsistencies in two of the relations which "should" have been predictable; therefore a unique higher-

ordered metric scale could not be constructed for him. At present, the nature of inconsistencies is being studied. One of the leads, suggested by Robert Radlow, is that inconsistencies are likely to occur in relations which involve equal-appearing intervals or combinations of intervals. The findings on the inconsistent subject seem to support this explanation. On the average, the time required to obtain a person's higher-ordered metric scale for five entities is twenty minutes.

REFERENCES

1. Birkhoff, G. Lattice theory. (Revised Ed.) New York: American Mathematical Society, Colloquium Publications, 1948, XXV.
2. Coombs, C. H. Psychological scaling without a unit of measurement. *Psychol. Rev.*, 1950, **57**, 145–158.
3. Coombs, C. H. Mathematical models in psychological scaling. *J. Amer. statist. Assn.*, 1951, **46**, 480–489.
4. Coombs, C. H. A theory of psychological scaling. Ann Arbor: Univ. of Michigan Engineering Research Institute, Bulletin, 1952, **34**.
5. Coombs, C. H. Theory and methods of social measurement. In L. Festinger and D. Katz (Eds.): Research methods in the behavioral sciences. New York: Dryden, 1953. Pp. 471–535.
6. Davidson, D., Siegel, S., and Suppes, P. Some experiments and related theory in the measurement of utility and subjective probability. Report No. 4, Stanford Value Theory Project, 1955.
7. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.
8. Ramsey, F. P. Truth and probability. In: The foundations of mathematics and other logical essays. London: K. Paul, Trench, Trubner, 1931.
9. von Neumann, J., and Morgenstern, O. Theory of games and economic behavior. (2nd Ed.) Princeton: Princeton Univ. Press, 1947.

Manuscript received 2/7/55

Revised manuscript received 4/4/55

BOOK REVIEWS

ISADORE BLUMEN, MARVIN KOGAN, and PHILIP J. MCCARTHY. *The Industrial Mobility of Labor as a Probability Process.* Ithaca: New York State School of Industrial and Labor Relations, Cornell University, 1955, xii + 163 pp. $3.00 paper, $4.00 cloth. (*Cornell Studies in Industrial and Labor Relations*, Vol. 6.)

This is an investigation of how the theory of Markov chains can be used, or adapted, for a description of the observed movements of the labor force within the United States. The phenomenon of labor movements has no direct connection with psychometrics, but the theory of Markov chains appears to a useful tool for the construction of models of learning.* Psychometricians may therefore find interest in the book from a purely methodological point of view. In fact, probably never before has so huge a statistical material been treated numerically by probabilistic models of the Markovian variety. The analysis is made expertly and described carefully to the last detail. The problems of handling the data, of statistical estimation, and of comparing theory with observations appear thus with great clarity. The shortcomings of the method are discussed with commendable frankness.

Consider, say, nine categories of employment and add the tenth category entitled "not covered by the preceding ones." Fix an arbitrary time unit and consider the workers who at time t are in category i. At time $t + 1$ a fraction f_{ij} of them will be found in category j (where $f_{i1} + \cdots + f_{i10} = 1$). In a stable community these frequencies will be (approximately) independent of t.

Denote by F_1 the ten by ten matrix with elements f_{ij}, and similarly by $F_2, F_3, \cdots$ the analogous matrices for an observational period of length 2, 3, $\cdots$. The simple Markov chain model assumes that the transitions from category i to other categories constitute a random choice which is in no way affected by the past history (for example, of the time spent in category i). If this were the case, the matrix F_1 should be nearly equal to the matrix P of the theoretical transition probabilities, and $F_2, F_3, \cdots$ should be close to the powers $P^2, P^3, \cdots$ of the matrix P.

The authors show how to estimate P and find that in actual practice the diagonal elements of $F_2, F_3, \cdots$ consistently and significantly exceed the corresponding elements of $P^2, P^3, \cdots$. This indicates that (contrary to the assumption underlying a Markov process) a prolonged employment in a category decreases the probability of a move into another category. Accordingly the authors refine the model by assuming that the entire population consists of two strata, the "stayers" and the "movers." The stayers never move whereas the movers are subject to a random process of the type just described. If the relative sizes of the two strata are p and $q = 1 - p$, respectively, the present model predicts that $F_n = pI + qP^n$, where I is an identity matrix. However, this model actually overestimates the diagonal elements of F_n. The present model is a special case of an exceedingly flexible and useful model to which the authors call attention. Instead of assuming that the stayers never move, we may assume that each of the movements in each stratum are subject to a Markov process as described above with matrices P and Q, respectively. Now we should have $F_n = pP^n + qQ^n$. Instead of two strata one may consider a larger number of strata, thus attaining higher accuracies. (Further modifications are indicated in the last chapter.) Models of this type could be useful for learning theory even more than for labor movements (where probably the after-effect of the past history is so pronounced that higher-order Markov chains must be introduced).

William Feller

Princeton University

*Cf. R. R. Bush and F. Mosteller, *Stochastic Models for Learning.* New York: Wiley, 1955.

PAUL E. MEEHL. *Clinical and Statistical Prediction.* Minneapolis, Minnesota: University of Minnesota Press, 1955, pp. x + 149.

In a little study I made recently of the interest and value structure of psychologists, the first principal component separates the laboratory from the clinic, the second separates the global-verbal from the analytic-psychometric approach to personality. It is hardly surprising, therefore, that the clinician and the statistician see things in different ways and have difficulty in communicating with one another. Meehl has attempted, in his present monograph, to achieve some reconciliation of these views and to build common ground between these groups.

Speaking as a psychometrician-statistician, I feel he has been quite successful. At least, there was little that I felt inclined to quarrel about in his presentation, and I feel I have a more sympathetic understanding of the clinician's activities as a result of my reading. I have not obtained the reactions of any thorough-going clinicians to find whether they felt equally satisfied.

There is not, as Meehl points out, one single clear issue as between the clinically oriented and the statistically oriented psychologist. Rather, there are a series of sub-problems. Thus, one issue concerns the value of psychometric as compared with non-psychometric *data*. A second concerns the efficiency of mechanical and non-mechanical ways of *combining* either type of data for purposes of prediction. One important distinction is between *structural statistics*, which aspire to provide a framework of constructs to describe the nature of the individual, and *validation statistics*, which undertake only to indicate the degree of association between prediction (whether stemming from scores or from clinical insights) and the course of subsequent events. A second, and perhaps the most central, contrast that Meehl makes is between the *context of discovery* and the *context of justification*.

A major part of the monograph is devoted to reviewing existing studies comparing the mechanical and non-mechanical modes of combining data for predictive purposes. Though the studies have many limitations, their general trend seems to definitely favor mechanical modes of combining, when the criterion to be predicted is some pre-established set of socially defined categories. These are such categories as level of academic grades, progress in recovery under therapy, or lapsing from grace during parole.

The primary predictive function for the clinician, as Meehl sees it, lies within the context of discovery in relation to the individual. That is, the distinctive contribution of the skilled clinician is that he can create the hypotheses that relate and apply general psychological principles to the uniqueness and complexity of the individual case. With what regularity these hypotheses are supported by later events remains a statistical problem. The need for tests of the accuracy of this hypothesizing is as great as the need to test the accuracy of predictions resulting from any psychometric device.

Teachers College, Columbia University

Robert L. Thorndike

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

THE RELATION BETWEEN INFORMATION AND VARIANCE ANALYSES*

W. R. GARNER

THE JOHNS HOPKINS UNIVERSITY

AND

WILLIAM J. MCGILL

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Analysis of variance and uncertainty analysis are analogous techniques for partitioning variability. In both analyses negative interaction terms due to negative covariance terms that appear when non-orthogonal predictor variables are allowed may occur. Uncertainties can be estimated directly from variances if the form of distribution is assumed. The decision as to which of the techniques to use depends partly on the properties of the criterion variable. Only uncertainty analysis may be used with a non-metric criterion. Since uncertainties are dimensionless (using no metric), however, uncertainty analysis has a generality which may make it useful even when variances can be computed.

I. *Introduction*

Shannon (5) has defined amount of information by the formula

$$H(y) = -\sum_{k=1}^{r} p(k) \log_2 p(k), \tag{1}$$

where y has r discrete values, and $p(k)$ is a probability distribution defined over y. In communication theory, y is considered a source of signals, and the measure H represents the average number of binary digits required to code or store one of the signals. A broader interpretation, however, makes H a parameter which measures the non-metric variability of any probability distribution. H has a value of zero when the probability is concentrated in a single category and is maximum when the probability is uniformly distributed over all categories.

Psychologists have been attracted by the non-metric character of this measure and the obvious application to situations where variances cannot be computed. Since this use of the measure is concerned only with its statistical properties and not with its interpretation in communication theory, we

*The work of the senior author was supported by Contract N5ori-166, Task Order 1, between the U.S. Office of Naval Research and The Johns Hopkins University. This is Report No. 166-I-192, Project Designation No. NR 145-089, under that contract.

Bureau Ednl. ... Research
... ING COLLEGE
Dated ...
Acc. No ...

shall use the more general term uncertainty, U, to refer to the measure. We shall show that uncertainty has many of the properties of variance and can be partitioned into components as variance can.

II. *The Analysis Problem*

The relations discussed apply when a criterion is predicted from one or more predictors. The development will be presented for the three-variable case, where the problem is to determine to what extent values of the criterion variable can be predicted from two predictor variables.

Our notation is as follows: The criterion variable y can assume any value y_k. The two predictor variables w and x can assume values w_i or x_j. We assume that all three variables are categorized in order that the formulas for uncertainty and variance analysis may have equivalent notations. This assumption does not limit any of the principles demonstrated.

In the three-dimensional matrix, n_{ijk} refers to the number of cases in a single cell; $n_{ij.}$ refers to the total number of cases having the ith value of w and the jth value of x; and $n_{i..}$ refers to the total number of cases having the ith value of w. Similar subscripts indicate other combinations of the three variables; n with no subscript indicates the total number of cases in the matrix. In analysis of variance formulas, $\bar{y}$ indicates a mean value, and the subscript notation just illustrated is used for mean values of the subclassifications.

III. *The Nature of Uncertainty Analysis*

Analysis of variance can be considered as two separate processes. First, the variance of the criterion variable is partitioned into its several identifiable components—components which add up to the total variance. This process is a simple descriptive one; there are no probability assumptions involved in its use. One describes the components of a total variance, making no assumptions about the distributions from which the data are drawn. The second process, which is not a necessary consequence of the first, involves using these partitioned components to obtain estimates of population variances and to make inferences about the parent population. For this process, the actual data provide sample estimates of population distributions; here assumptions about the population distributions become critical.

Uncertainty analysis likewise has both processes. The first process is purely descriptive: it is intended to allow the partitioning of the uncertainty of the criterion variable $U(y)$ into components. Since this process is entirely descriptive, there are no underlying probability assumptions. All that is required for its use is that a data matrix of the type described above is available. The primary purpose of this paper is to demonstrate the nature of uncertainty partitioning and to compare it to variance partitioning. This process is illustrated and explained in Table 1. The results of uncertainty

partitioning specify sources and magnitudes of variabilities as well as amount of categorical discrimination available. These uses are explained more fully by Garner and Hake (1) and by McGill (3).

The n_{ijk} can be considered as sample estimates of $p(i, j, k)$; we can use these sample estimates to test various hypotheses about the parent distribution. For example, suppose we wish to test the hypothesis that both predictors are independent of the criterion, i.e.,

$$p(i, j, k) = p(i, j)p(k). \tag{2}$$

It can be shown (3, 4) by using the likelihood ratio, that when hypothesis (2) is true, $[1.3863\ nU(y\colon w, x)]$ is distributed approximately as chi square. Independent tests can be constructed in the same way for each of the predictors separately as well as for the interaction between predictors. The approximation to chi square is of the same order as the familiar chi-square contingency test so that, in effect, uncertainty analysis is analysis of contingency chi square. Miller and Madow (4) discuss this aspect of uncertainty analysis more thoroughly.

IV. *The Orthogonal Case*

Usually in analysis of variance and in uncertainty analysis, the experimenter tries to set up orthogonal predictions. Orthogonality is defined as zero association between the predictors. This requirement is met when the cell frequencies in the matrix of the $n_{ij.}$ can be predicted correctly from the row and column marginal frequencies, i.e., when

$$n_{ij.} = \frac{n_{i..}n_{.j.}}{n}. \tag{3}$$

Uncertainty Analysis

The partitioning of $U(y)$ in uncertainty analysis is illustrated by

$$U(y) = U(y\colon w, x) + U_{wx}(y), \tag{4}$$

where the uncertainty measures have the definitions given in Table 1. The second term on the right-hand side of (4) is the error uncertainty, i.e., the amount of uncertainty in the criterion y remaining after the predictable uncertainty has been eliminated. The first term on the right-hand side of (4) is the predictable uncertainty; it in turn can be partitioned into components

$$U(y\colon w, x) = U(y\colon w) + U(y\colon x) + U(y\colon \overline{wx}). \tag{5}$$

These terms are also defined in Table 1. A feature of uncertainty analysis is the interaction term $U(y\colon \bar{w}\bar{x})$. This is the uncertainty in y predictable from unique combinations of w and x.

Equation (5) describes a process that is identical in form with the

TABLE 1

Symbols, Formulas, and Definitions Used in Three-variable Uncertainty Analysis

The criterion variable, y, is assumed to be non-metric. The predictor variables, w and x, are categorized and may or may not be metric variables.

	Symbol	Formula	Definition
(1)	$U(y)$	$-\sum_k \left[\frac{n_{..k}}{n}\right] \log_2 \left[\frac{n_{..k}}{n}\right]$	*Total uncertainty:* The amount of uncertainty in the criterion variable, y.
(2)	$U_w(y)$	$-\sum_i \left[\frac{n_{i..}}{n}\right] \sum_k \left[\frac{n_{i.k}}{n_{i..}}\right] \log_2 \left[\frac{n_{i.k}}{n_{i..}}\right]$	*Conditional uncertainty:* The amount of uncertainty in y when one predictor variable, w, is held constant.
(3)	$U_{wx}(y)$	$-\sum_{i,j} \left[\frac{n_{ij.}}{n}\right] \sum_k \left[\frac{n_{ijk}}{n_{ij.}}\right] \log_2 \left[\frac{n_{ijk}}{n_{ij.}}\right]$	*Error uncertainty:* The amount of uncertainty remaining when both predictor variables, w and x, are held constant.
(4)	$U(y:w)$	$U(y) - U_w(y)$	*Contingent uncertainty:* The uncertainty in y due to the predictor variable, w.
(5)	$U_w(y:x)$	$U_w(y) - U_{wx}(y)$	*Partial contingent uncertainty:* The uncertainty in y due to the predictor variable, x, when the predictor variable, w, is held constant.
(6)	$U(y:w,x)$	$U(y) - U_{wx}(y)$	*Multiple contingent uncertainty:* The uncertainty in y due to the joint influence of the predictor variables, w and x.
(7)	$U(y:\overline{wx})$	$-U(y) + U_w(y) + U_x(y) - U_{wx}(y)$, or $U_w(y:x) - U(y:x)$	*Interaction uncertainty:* The uncertainty in y due to unique combinations of the predictor variables, w and x.

TABLE 2

Symbols, Formulas, and Definitions Used in Three-variable Analysis of Variance

The criterion variable, y, is assumed to be metric. The predictor variables, w and x, are categorized but not necessarily metric.

	Symbol	Formula	Definition
(1)	$V(y)$	$\sum_k \left[\frac{n_{..k}}{n}\right] (y_k - \bar{y})^2$	*Total variance:* The variance of the criterion variable, y.
(2)	$V_w(y)$	$\sum_i \left[\frac{n_{i..}}{n}\right] \sum_k \left[\frac{n_{i.k}}{n_{i..}}\right] (y_k - \bar{y}_{i..})^2$	*Conditional variance:* The variance of y when one predictor variable, w, is held constant.
(3)	$V_{wx}(y)$	$\sum_{i,j} \left[\frac{n_{ij.}}{n}\right] \sum_k \left[\frac{n_{ijk}}{n_{ij.}}\right] (y_k - \bar{y}_{ij.})^2$	*Error variance:* The variance remaining in y when both predictor variables, w and x, are held constant.
(4)	$V(y:w)$	$V(y) - V_w(y)$	*Main effect:* The variance of y due to the predictor variable, w.
(5)	$V_w(y:x)$	$V_w(y) - V_{wx}(y)$	*Partial main effect:* The variance of y due to the predictor variable x, when the predictor variable, w, is held constant.
(6)	$V(y:w,x)$	$V(y) - V_{wx}(y)$	*Total predictable variance:* The variance of y due to the joint influence of the predictor variables, w and x.
(7)	$V(y:\overline{wx})$	$-V(y) + V_w(y) + V_x(y) - V_{wx}(y)$, or $V_w(y:x) - V(y:x)$	*Interaction variance:* The variance of y due to unique combinations of the predictor variables, w and x.

partitioning of variance in analysis of variance; in the orthogonal case the interaction uncertainty can be interpreted by analogy with interaction variance. This is true despite the fact that interaction uncertainties are sometimes negative (3). This problem will be discussed in detail in Section V.

Analysis of Variance

Uncertainty analysis is generally appropriate when the criterion variable y is a categorical variable, i.e., one allowing only nominal scale values (cf. 6). The predictor variables may be categorical, or they may be metric variables which are categorized for purposes of analysis. If the criterion is a true metric variable, i.e., one having at least the properties of an interval scale, we can compute variances and perform analysis of variance. The predictor variables must be categorized in any simple form of the analysis of variance.

Equations describing analysis of variance are essentially identical to those of uncertainty analysis. The defining equations are given in Table 2; except for the fact that variances are computed from squared deviations, whereas uncertainties are computed from log-probabilities, the equations are identical to those in Table 1. The partition of the variance of the criterion can be written:

$$V(y) = V(y: w, x) + V_{wx}(y). \tag{6}$$

Again the two parts on the right-hand side of the equation are the predictable and the error components of the total variance. The predictable variance can be broken down as before:

$$V(y: w, x) = V(y: w) + V(y: x) + V(y: \overline{wx}). \tag{7}$$

The terms in (7) are explained in detail in Table 2.

Normally the analysis of variance in (7) is called double classification; the variances are generally identified in terms of the two predictors. This shorthand procedure is convenient for most purposes. However, it obscures the fact that the data array is three-dimensional. The analysis is identical to the one treated in uncertainty analysis in every respect, except that in the analysis of variance the criterion variable has a metric, whereas it does not in uncertainty analysis.

V. *The Non-Orthogonal Case*

In Section IV it was mentioned that the interaction term in uncertainty analysis can assume negative values under certain conditions. It is equally true that the interaction term in analysis of variance can be negative, if it is defined as in Table 2. The negative interaction term is due to non-orthogonality and can be thought of as due to a negative covariance term that may attenuate or exceed the positive interaction effect.

Uncertainty Analysis

It is not difficult to show that the interaction uncertainty in (5) can be written

$$U(y: \overline{wx}) = U_v(w: x) - U(w: x). \tag{8}$$

This form of the interaction term shows at once that interaction cannot be negative with orthogonal predictors since orthogonality requires that $U(w: x) = 0$.

In the non-orthogonal case, however, $U(w: x)$ will be greater than zero. With certain combinations of cell frequencies, the contingent uncertainty between x and w can be larger than the partial contingent uncertainty—resulting in negative interaction. A simple illustration of this principle is provided when each value of w is paired uniquely with each value of x. Now $U(w: x)$ is as large as it can be. Furthermore, $U_v(w: x)$ cannot be greater than $U(w: x)$ since $U(w: x)$ is the maximum contingent uncertainty that can be obtained from a contingency table involving w and x. Equation (8) shows that the interaction will never be greater than zero. An identical result is obtained in the variance analysis when the predictors are completely confounded.

Analysis of Variance

It is usually assumed that the components of the total variance in analysis of variance must be positive. This is true only in the orthogonal case; if an analysis of variance is carried out with a non-orthogonal experimental design, using the equations given in Table 2, negative interaction terms can occur.

To show how this happens, we now analyze the components of the interaction variance for the general case. The equation is

$$V(y: \overline{wx}) = \frac{1}{n} \sum_{i,j} n_{ij.}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y})^2 - \frac{2}{n} \sum_{i,j} \left(n_{ij.} - \frac{n_{i..}n_{.j.}}{n}\right)(\bar{y}_{i..} - \bar{y})(\bar{y}_{.j.} - \bar{y}). \tag{9}$$

It can be seen that the interaction variance is composed of two parts: the first part is essentially the interaction variance in the orthogonal case; the second part is a negative covariance term. This term must be zero in the orthogonal case [see equation (3)], but in the non-orthogonal case it cannot be ignored. The redundancy introduced by non-orthogonality is illustrated clearly in multiple regression. No interaction term is permitted, but a correction for non-orthogonality must be introduced whenever the predictor variables are correlated (cf. 2).

VI. *Effects of Non-Orthogonality*

Our discussion of non-orthogonality shows that it is best to design experiments with orthogonal predictor variables. The analysis is simplified, and the uninterpretable interaction components are eliminated.

Clearly the covariance in (9) is not just part of the interaction variance. In fact, when predictors are non-orthogonal, the concept of interaction is almost meaningless. For example, consider an analysis of variance in which the predictors w and x are completely confounded. The two main-effect variances and the interaction will all be identical. The covariance term in (9) must be large enough to cancel out two of these variances, but we do not know which two of the variances should be cancelled out. In a sense, the covariance term is a correction factor which must be applied to the entire set of variances. Thus, a covariance term (whether or not it is large enough to produce a negative interaction) renders an exact interpretation of the component variances impossible.

The multiple contingent uncertainty or the total predictable variance can be computed directly as shown in the defining equations in Tables 1 and 2. The negative covariance term is included; there is no over-estimation of the total predictable variance or uncertainty. However, the interpretation of results should be made only in terms of combinations of the two predictors —no valid statements can be made about them independently.

Sometimes it is impossible to obtain orthogonal predictor variables, particularly when there are more than two. In time series successive events are usually not orthogonally related because no independent control of these events is possible. If the time series has serial dependencies, preceding events cannot be orthogonal. Consequently, the total predictability of events in a time series cannot in general be computed by adding up the separate predictabilities obtained from preceding events displayed by one or more units in the time series.

VII. *Estimation of Uncertainties from Variances*

It is clear that uncertainty analysis and analysis of variance are analogous analytic techniques. In fact, variances may be used to estimate uncertainties if we assume that y is normally distributed.

Shannon (5) has shown that the uncertainty of a normal distribution can be specified as

$$\text{est } U(y) = \tfrac{1}{2} \log_2 2\pi e V(y) - \log_2 m, \tag{10}$$

where est $U(y)$ is the estimated total uncertainty of the criterion variable on the assumption of a normal distribution of values of y_k, and where m is the width of the category interval on the y continuum.

We can write similar equations for any of the variances obtained in

analysis of variance. For example,

$$\text{est } U_{wx}(y) = \tfrac{1}{2} \log_2 2\pi e V_{wx}(y) - \log_2 m \tag{11}$$

is the error uncertainty estimated from error variance. From definition (6) in Table 1, and from equations (10) and (11), we can write

$$\text{est } U(y: w, x) = \tfrac{1}{2} \log_2 [V(y)/V_{wx}(y)]. \tag{12}$$

Thus, it is relatively simple to estimate the multiple contingent uncertainty from the appropriate variances. The expression on the right-hand side of this equation is reminiscent of the multiple correlation ratio (η). We can, in fact, write

$$\text{est } U(y: w, x) = -\tfrac{1}{2} \log_2 [1 - \eta^2(y: w, x)]. \tag{12–A}$$

Estimated uncertainties have the properties of additivity observed in computed uncertainties. Consequently, the expression on the right-hand side of (12) can be partitioned into three components, each of which is based on its equivalent variances as follows:

$$\text{est } U(y: w) = \tfrac{1}{2} \log_2 [V(y)/V_w(y)], \tag{13}$$

$$\text{est } U(y: x) = \tfrac{1}{2} \log_2 [V(y)/V_x(y)], \tag{14}$$

$$\text{est } U(y: \overline{wx}) = \tfrac{1}{2} \log_2 \{[V_w(y) \cdot V_x(y)]/[V(y) \cdot V_{wx}(y)]\}. \tag{15}$$

These estimating equations point out some of the differences between uncertainty and variance. If (15) is used to estimate the interaction uncertainty when the interaction variance is zero, cases can be found in which the estimated interaction uncertainty (and the computed interaction uncertainty) will not be zero. Converse cases (i.e., zero uncertainty interactions with finite variance interactions) can also be found. These apparent contradictions are due to the fact that variances and uncertainties, while analogous, do not measure exactly the same characteristics of probability distributions. Uncertainty analysis depends on the number of categories occupied by a distribution. Variance analysis depends on the weights or values attached to these categories.

VIII. *Application of the Measures*

We have now shown that uncertainty analysis and analysis of variance are equivalent in many respects; the question naturally arises as to when one should be used in preference to the other. This decision depends on the properties of the data and the assumptions the experimenter is willing to make. If the criterion variable y has only the properties of a nominal or ordinal scale, then only uncertainty analysis is permissible. Uncertainty analysis has the greater generality and requires no assumptions about metric properties of the criterion.

On the other hand, uncertainty analysis does not give any information about the metric if it exists. If the criterion variable is metric with at least the properties of an interval scale, then analysis of variance must be used to retain information about the metric. The variance measure in retaining the metric sacrifices generality since the variances obtained from one experiment are not directly comparable to those obtained from another. Thus, the fact that the uncertainty measure is dimensionless gives it a generality which allows direct comparison of experimental results which differ in their metric.

To summarize, the measures are similar in many respects, but they are not identical. The uncertainty measure has greater generality and the advantages of generality. The variance measure is more specific but retains information about the metric. The decision as to which to use depends not only upon the properties of the criterion variable but also upon the gain expected from being more sensitive instead of more general. In many applications it is reasonable to use both measures and compare them.

REFERENCES

1. Garner, W. R. and Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, 58, 446-459.
2. Garner, W. R. and McGill, W. J. Relation between uncertainty, variance, and correlation analyses. Mimeographed report, Department of Psychology, The Johns Hopkins University, Baltimore, Maryland.
3. McGill, W. J. Multivariate information transmission. *Psychometrika*, 1954, 19, 97-116.
4. Miller, G. A. and Madow, W. G. On the maximum likelihood estimate of the Shannon-Wiener measure of information. Report AFCRC-TR-54-75. Operational Applications Laboratory, ARDC, August 1954.
5. Shannon, C. E. A mathematical theory of communication. *Bell syst. tech. J.*, 1948, 27, 379-423, 623-656.
6. Stevens, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley, 1951, pp. 1-49.

Manuscript received 10/9/54

Revised manuscript received 9/23/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

MEASUREMENT OF SUBJECTIVE VALUES

HAROLD GULLIKSEN

PRINCETON UNIVERSITY
AND
EDUCATIONAL TESTING SERVICE

Four different value laws are developed and tested by using them to predict the scale values of composite stimuli from the scale values of their components. These four laws are: an additive law, a square-root law, a logarithmic, and a negative exponential law. They are tried out on a set of food preferences by means of Pearson's Method of False Position. The negative exponential law of diminishing returns gave the best fit to the data but was not markedly better than any of the other laws.

The purpose of this study is to show that laws relating subjective value to amount of commodity may be studied by an extension of the usual psychophysical scaling methods. Four different value laws will be developed and tested with a set of experimental data on food preferences.

The psychophysical scaling procedures, such as paired comparisons, for example, may be used to distinguish between various types of laws expressing value increase as a function of increase in amount of the commodity. These procedures are applicable even when no physical measurement of the amount of the commodity can be made, and when the scaling procedure necessitates measuring from an arbitrary origin.

Testing each of these laws of value increase also involves a corresponding determination of an origin, or point of zero value. Various psychophysical methods of determining a zero point have been presented in the literature. An additive law of value increase has been used for this purpose (cf. **1**, **4**, and **7**). It will be shown here that other laws of value increase may also be used to determine an origin or point of zero value.

The data necessary to test these formulations are obtained by using a preference schedule like that illustrated in Table 1. The subject is asked the usual paired comparison question, "Which do you prefer, i or j?" However, in addition to the single stimuli i and j *composite* stimuli of the form (i and j) or (g and h) are used. The subject is asked questions of the form, "Do you prefer (i and j) or g?" as well as, "Do you prefer (i and j) or (g and h)?" (**15**).

Do these different stimuli, designated by i, j, g, h, (i and j), (g and h), etc., behave as if they are different amounts of some commodity x whose subjective value v is given by some function, say $v = f(x)$? For any given function, the experimental device of utilizing the *composite* stimuli of the

type (i and j) may be indicated by writing

$$v_i = f(x_i); \qquad v_j = f(x_j); \quad \text{and} \quad v_{ij} = f(x_i + x_j). \tag{A}$$

For example, if the items i, j, and (i and j) were bundles of dollars, such that the number of dollars in each bundle were known to the subject, but not to the experimenter, then the experimenter could raise the question asked here. Do these different bundles and their combinations behave as if they were different packages containing different numbers of dollars x, which were known to the subject and were related to the subjective value v by some function $v = f(x)$?

It is of interest to note that since

$$x_i = f^{-1}(v_i); \qquad x_j = f^{-1}(v_j); \quad \text{and} \quad x_i + x_j = f^{-1}(v_{ij}), \tag{B}$$

we have

$$f^{-1}(v_{ij}) = f^{-1}(v_i) + f^{-1}(v_j). \tag{C}$$

From the viewpoint of psychological scaling, we do not obtain v's, but instead obtain s's which differ from the v's by a constant, so we may write

$$f^{-1}(s_{ij} - c) = f^{-1}(s_i - c) + f^{-1}(s_j - c). \tag{D}$$

In other words, if one can find any function, designated f^{-1}, such that an additive relationship holds for the proper scale values as indicated then one possible solution for the relationship between v and x is the inverse f of this function f^{-1}. From this point of view the commodity amounts are simply defined by $x_i = f^{-1}(v_i)$.

If another function g is found such that

$$g^{-1}(v_{ij}) = g^{-1}(v_i) + g^{-1}(v_j), \tag{E}$$

we may regard it as defining another commodity amount $y_i = g^{-1}(v_i)$. It is of interest to ask about the relationship between x_i and y_i or between g^{-1} and f^{-1}. Substituting (A) in (E),

$$g^{-1}f(x_i + x_j) = g^{-1}f(x_i) + g^{-1}f(x_j). \tag{F}$$

A theorem in functional equations states that if

$$F(X + Y) = F(X) + F(Y), \tag{G}$$

then

$$F(X) = aX, \tag{H}$$

where a is a constant coefficient. This theorem was proved for continuous functions by Cauchy (2). The discontinuous solution for (G) is discussed by Hamel (6) and Sierpiński (11). Applying solution (H) to (F) we have

$$g^{-1}f(x_i) = a(x_i).$$

Utilizing (B) gives

$$g^{-1}(v_i) = af^{-1}(v_i) \quad \text{or} \quad y = ax.$$

In other words, the "different" value laws found in this way would differ only in the unit of measurement used for the commodity. It should be noted that this statement holds only if equations (B) and (E) hold without error for values of v_i, v_j, and v_{ij}.

In order to test different value laws, the theoretical formulation will need to show the relationship between the values v_i and v_j for the *component* stimuli and the value v_{ij} of the *composite* stimulus. Let us see what different laws of value increase imply with respect to this relationship.

Four Laws of Value Increase

Logarithmic Law

If it is assumed that the rate of change in value v is inversely proportional to the amount of the commodity x, we have the differential equation

$$dv/dx = k/x. \tag{1}$$

Integrating this equation and setting $x = x_0$ when $v = 0$, we have

$$v = k \log (x/x_0). \tag{2}$$

This derivation of a logarithmic law of value increase was given by Thurstone (13).

What does this logarithmic value law imply with respect to the relationship between the values of the *component* and the *composite* stimuli? In order to determine this, we note that

$$\begin{aligned} v_{ij} &= k \log [(x_i/x_0) + (x_j/x_0)], \\ v_i &= k \log (x_i/x_0), \end{aligned} \tag{3}$$

and

$$v_j = k \log (x_j/x_0).$$

Eliminating x_i and x_j among these equations enables us to find an expression for the composite value v_{ij} in terms of the values v_i and v_j of the components. Thus, we find that

$$e^{v_{ij}/k} = e^{v_i/k} + e^{v_j/k}. \tag{4}$$

These equations do not require the measurement of amount of commodity x, but require only a set of value measurements v. However, the usual scaling data give neither the v's nor the x's but only scale values, s, which differ from v by a constant. Let C be the scale value for which v equals zero.

By making the substitutions

$$v_i = s_i - C,$$
$$v_j = s_j - C, \tag{5}$$

and

$$v_{ij} = s_{ij} - C,$$

let us determine the type of relationship among the scale values of the component stimuli and the composite stimuli that is implied by the logarithmic law of value increase. Substituting and simplifying gives

$$e^{s_{ij}/k} = e^{s_i/k} + e^{s_j/k}. \tag{6}$$

This equation gives the interrelationships among the experimentally determinable scale values which are implied by the logarithmic law of value increase.

There are some interesting and disconcerting things about this law. It contains the parameter k, which is involved in such a way that it cannot readily be solved for explicitly. Thus, it remains as an annoying trial parameter in any attempts to verify this equation. It is also interesting to note that the equation does not contain C at all. In other words, for the logarithmic law of value increase the additive constant C cannot be determined. *Any* additive constant is consistent with the relationships among the component and composite scale values. Likewise, any value of x_0 is consistent with these relationships, since for this law, x_0 functions as a unit of measurement for x.

Square-Root Law

If it is assumed that the rate of change in value v is inversely proportional to the value level already attained, we may write the differential equation

$$dv/dx = k/2v. \tag{7}$$

Integrating and setting $x = x_0$ when $v = 0$ we have

$$v = \sqrt{k(x - x_0)}. \tag{8}$$

This derivation of a square-root law of value increase was given by Thurstone (13).

To determine the implications of this square-root law with respect to the relationship between the values of the *component* and *composite* stimuli, we note that

$$v_i = \sqrt{k(x_i - x_0)}$$
$$v_j = \sqrt{k(x_j - x_0)} \tag{9}$$
$$v_{ij} = \sqrt{k(x_i + x_j - x_0)}.$$

Eliminating x_i and x_j among these equations enables us to find an expression for the composite value v_{ij} in terms of the values v_i and v_j of the components. Thus, we find that

$$v_{ij}^2 = v_i^2 + v_j^2 + kx_0 \,. \tag{10}$$

In order to determine the relationship among the scale values implied by this equation, we substitute (5) in (10), giving

$$s_{ij}^2 - s_i^2 - s_j^2 = 2C(s_{ij} - s_i - s_j) + C^2 + kx_0 \,. \tag{11}$$

This equation gives the interrelationship among the experimentally determinable scale values which is implied by the square-root law of value increase. The plot of the quantity on the left side of the equation against the quantity in parentheses enables one to find C from the slope of the line, and also to find kx_0 by subtracting the square of half the slope from the intercept. Separate values for k and x_0 cannot be determined.

Negative Exponential Law

If it is assumed that the rate of change in value v is directly proportional to the difference between the value level already reached and an asymptotic value level A, we have the differential equation

$$dv/dx = k(A - v). \tag{12}$$

Integrating and setting $x = x_0$ when $v = 0$ gives

$$v = A - Ae^{kx_0}e^{-kx}. \tag{13}$$

This is the familiar negative exponential law of diminishing returns used in economics (see, for example, **8** and **12**). It has also been suggested by a number of writers as an equation of the learning curve (see illustrations cited in **3**). If we apply it to the component stimuli "i" and "j" and also to the composite "i and j", we have

$$\begin{aligned} v_i &= A - Ae^{kx_0}e^{-kx_i}, \\ v_j &= A - Ae^{kx_0}e^{-kx_j}, \\ v_{ij} &= A - Ae^{kx_0}e^{-k(x_i+x_j)}. \end{aligned} \tag{14}$$

Eliminating x_i and x_j to find v_{ij} as a function of v_i and v_j gives

$$v_{ij} = A - (1/A)e^{-kx_0}(A - v_i)(A - v_j). \tag{15}$$

If v_i is plotted against v_{ij} for a given value of v_j, the result is a straight line. If this plot is made for each of the values of v_j, the result is a family of straight lines. We also note that if either v_i or v_j is equal to A, then the right-hand term of the equation vanishes, giving $v_{ij} = A$. Thus, the indicated plots constitute a *pencil* of straight lines intersecting in the point (A, A). Again

we may substitute for the v's in terms of the s's (equations 5), obtaining

$$s_{ij} = A + C - (1/A)e^{-kx_0}(A + C - s_i)(A + C - s_j). \tag{16}$$

Again, if s_i is plotted against s_{ij} for a given value of s_j, the result is a straight line. If such a plot is made in turn for each of the possible values of s_j the result is a *pencil* of straight lines intersecting in the point $(A + C, A + C)$. If $x_0 = 0$ this series of straight lines may be used to give the values of A and of C.

Linear Law

If we assume that the rate of change of value is constant, we have the differential equation

$$dv/dx = k. \tag{17}$$

Integrating and setting $x = x_0$ when $v = 0$, we have

$$v = k(x - x_0). \tag{18}$$

If we solve for the interrelationships among the component and composite scale values implied by this equation, we find that

$$s_{ij} = s_i + s_j - C + kx_0 . \tag{19}$$

This relationship has been derived by Thurstone and utilized in an unpublished study (**15**). According to this law, the plot of the scale value of the composite against the sum of the scale values of the components should be linear with unit slope and intercept $kx_0 - C$.

The Food Preference Experiment

Let us now consider the type of data that is used for these value studies. Food preferences were studied.

Pairs of single items are presented such as *Beef* vs. *Pork*, and the subject is asked to indicate which he would choose, i or j. Then pairs of what we shall term *composite* items are presented. The subject is asked to choose between *Beef and Steak* vs. *Tongue and Lamb*. Also the choice is given between single and composite items, such as *Steak* vs. *Pork and Tongue*. Table 1 shows three typical items in the schedule.

The set of 5 component stimuli and 10 composite stimuli (making 15 stimuli in all) were presented in a paired comparison schedule to 92 college students in a psychology class. The directions stressed that for a composite such as *Beef and Steak*, *each* is an ordinary sized serving, and that if the composite were chosen, the person was to eat *both*, thus having twice as much as if only a single *component* item were chosen. This was done in order to give a better chance for diminishing returns to be manifested in the results. All choices involving a duplicate item were omitted, resulting in a matrix of

TABLE 1

Sample Questionnaire Items		
38.	☐	Roast Rib of Prime Beef
	☐	Roast Loin of Pork
39.	☐	Roast Rib of Prime Beef Sirloin Steak
	☐	Boiled Smoked Beef Tongue Loin Lamb Chop
40.	☐	Sirloin Steak
	☐	Roast Loin of Pork Boiled Smoked Beef Tongue

incomplete data. A least squares procedure was used (**5**) for scaling a paired comparisons matrix of incomplete data.

The scale values obtained range from .000 for tongue and pork, the least preferred item, on up through 1.043 for lamb, to 2.622 for the composite, beef and steak, which was the most preferred item. The complete set of scale values is shown in Table 2.

The fact that these 15 scale values give a good fit (**5**) to the 55 paired comparisons judgments shows that persons can make consistent judgments about preferences for composite stimuli along with single stimuli of the type used here. Thus, it is experimentally feasible to present in a single schedule comparisons of the (i vs. j), (i vs. g and h), and (i and j vs. g and h) types. Any set of concrete objects or even abstract concepts can be dealt with according to this pattern.

In Table 2 we notice that the value of tongue and lamb is higher than tongue alone, pork and lamb is higher than pork alone, beef and lamb is higher than beef alone, and lamb and steak is higher than steak alone. Thus, the value of any item is increased by forming a composite with lamb. The same holds true for beef and steak. Thus, lamb, beef, and steak are all positive values. However, now look at pork. The value of pork and steak is lower than the value of steak alone, of pork and beef lower than beef alone, of pork and lamb lower than lamb alone, and the value of the *composite* tongue and pork is lower than that of either of the components. Thus, from the purely ordinal characteristics of the scale, it seems clear that pork has a negative value. The same is true to an even greater extent for tongue. The zero point is between pork and lamb since pork and tongue are negative and lamb, beef, and steak are positive.

TABLE 2

Food Preference Experiment

Stimuli	Scale Values
Tongue and Pork	.000
Tongue	.137
Tongue and Lamb	.270
Pork	.541
Tongue and Beef	.830
Pork and Lamb	.928
Lamb	1.043
Tongue and Steak	1.088
Pork and Beef	1.448
Beef	1.746
Pork and Steak	1.780
Lamb and Beef	1.993
Steak	2.197
Lamb and Steak	2.324
Beef and Steak	2.622

Generalization to Include Positive and Negative Values

For this particular set of data there is thus clear evidence that some of the component values are positive and others are negative. The *linear* value law extends readily to include both positive and negative values and their various combinations.

The other laws, however, in their previously stated form do not give reasonable results for both positive and negative values. However, it is possible to make an appropriate extension by having four different rules, depending first on whether the components had the same or different signs, and second on whether the composite was positive or negative.

A reasonable interpretation including both negative and positive values and their combinations for the square-root law may be made by assuming $x_0 = 0$ and writing

$$\xi_{ij} = \frac{v_i + v_j}{|v_i + v_j|} \sqrt{\left| v_i^2 + \frac{v_i v_j}{|v_i v_j|} v_j^2 \right|} + C, \tag{20}$$

where $v_i = s_i - C$; $v_j = s_j - C$.

This formulation merely says that v_i^2 and v_j^2 are added if v_i and v_j are of the same sign and subtracted if v_i and v_j are of unlike signs. The square root of the sum is then given the sign of the larger value. If v_i and v_j are both positive, (20) is equivalent to (10) or (11). If v_i and v_j are both negative, (20) is equivalent to a reflection into the quadrant where s_i , s_j , and s_{ij} are each negative. If v_i and v_j are of opposite sign, (20) gives an interpretation that is consistent with the previous cases.

In order to state the logarithmic law for a series of either positive or negative values we have analogously with the square-root law:

$$\hat{s}_{ij} = \frac{v_i + v_j}{| v_i + v_j |} k \log \left| e^{|v_i|/k} + \frac{v_i v_j}{| v_i v_j |} e^{|v_j|/k} \right| + C, \tag{21}$$

where $v_i = s_i - C$; $v_j = s_j - C$.

Again this formula merely states that e is taken to be a positive power for either positive or negative values. The resulting powers are added if v_i and v_j are of the same sign and subtracted if v_i and v_j have different signs. The logarithm is then given the sign of the larger value. In contrast to (6), it is now possible to determine C for (21) since the *combination* of negative and positive values is involved.

The negative exponential law may be restated for the case in which both negative and positive values are involved. Let $x_0 = 0$; thus

$$\frac{A - v}{A} = e^{-kx} \qquad \text{(where } v = s - C\text{)}. \tag{22}$$

The formulas are easier to work with if we define B as the asymptote expressed in terms of the s-scale, just as C is defined:

$$B = A + C.$$

Let us assume that the positive (upper) asymptote B^+ may vary independently of the negative (lower) asymptote B^-. Then let us define

$$R_i = \frac{B^{\mp} - s_i^{\mp}}{| B^{\mp} - C |} = \frac{x_i}{| x_i |} e^{-k|x_i|}. \tag{23}$$

The superscript signs are used to indicate that the positive asymptote B^+ is used if $s_i > C$ (i.e., for s^+); the negative asymptote B^- is used if $s_i < C$ (i.e., for s^-). The R's thus defined are positive quantities if $s_i > C$, and negative if $s_i < C$. Note that as $| x_i |$ increases $| R_i |$ decreases. Then

$$\hat{s}_{ij} = B_i - (B_i - C) a R_j^a R_i , \tag{24}$$

where $a = R_i R_j / | R_i R_j |$ and $| R_j | > | R_i |$.

Equation (24) gives a set of computations expressed entirely in s-scale values by means of which an estimate $\hat{s}_{ij}$ of the scale value of a composite may be computed from the scale values of the components, assuming a negative exponential law of value increase.

Method of False Position

It was also found that the test equations previously developed were not sensitive enough to differentiate clearly between the different value laws. The method adopted was that of using the component stimuli to predict the composite value—selecting the parameters to minimize the sum of the squares of the differences between the observed and predicted scale values for the composite stimuli. For the linear additive rule, the solution is straightforward. For the others no explicit solution could be found, so the fit was made by a successive approximations procedure using Pearson's Method of False Position (**10**). A brief account of the method presented in matrix notation will be given here. This method is a general solution for linear or nonlinear equations presented by Karl Pearson. The problem may be stated as follows:

Given the k-parameter function

$$Y_i = f(m_1, m_2, \cdots, m_g, \cdots, m_k, x_i)$$

together with experimental observations of the paired values x_i, y_i ($i = 1, \cdots, n$), to determine the values of m_g so as to minimize $\sum_{i=1}^{n} (y_i - Y_i)^2$. Only two restrictive conditions are necessary:

(1) Given the value of the independent variable x_i and a set of arbitrary values m_{hg} for each of the parameters, it must be possible to compute (or to obtain by mechanical or other means) a corresponding set of values Y_{hi} for the dependent variable.

(2) The function $Y_{hi} = f(m_{h1}, m_{h2}, \cdots m_{hg}, \cdots m_{hk}, x_i)$ must be continuous and have continuous derivatives for slight changes in the values of the parameters in the neighborhood of the desired solution.

For a one-parameter function the Law of Mean Value may be stated as

$$\frac{dY_{pi}}{dm_{p0}} = \frac{Y_{hi} - Y_{0i}}{m_h - m_0},$$

or

$$Y_{hi} - Y_{0i} = \frac{dY_{pi}}{dm_{p0}}(m_h - m_0).$$

For a k-parameter function the general Law of the Mean for functions of several variables (cf. **9**, p. 121) gives

$$Y_{hi} - Y_{0i} = \sum_{g=1}^{k} \left[\frac{\partial Y_{pi}}{\partial m_{pg}} (m_{hg} - m_{0g}) \right],$$

where

i is an index for the observations of the dependent and independent variables ($i = 1, \cdots, n$),

g is an index for the parameters in the function ($g = 1, \cdots, k$),

0 indicates the initial guess for the values of the parameters,

h indicates subsequent guesses for the value of the parameters ($h = 1, \cdots, k$),

p indicates that the value of the partial derivative is taken at a point p on the curve between m_{0g} and m_{hg},

m_{hg} indicates the hth estimate for the parameter m_g.

Pearson's derivation is algebraic and quite voluminous. A much briefer statement in terms of matrices has been given by Gale Young in an unpublished note to the writer. This derivation is presented with acknowledgment to Dr. Young. We may put the derivation in matrix terminology as follows: Let

Y be a matrix of k rows and n columns with elements $Y_{hi} - Y_{0i}$ ($h = 1, \cdots, k$; $i = 1, \cdots, n$).

M be a square matrix with elements $m_{hg} - m_{0g}$, where ($g = 1, \cdots, k$; $h = 1, \cdots, k$).

F_p be a matrix of k rows and n columns with elements $\partial Y_{pi}/\partial m_{pg}$.

Then from the Law of Mean Value

$$Y = MF_p \quad \text{or} \quad M^{-1}Y = F_p .$$

Thus, we have a means of eliminating matrix F_p, for which it would be very difficult to find reasonable experimental values. In order for M^{-1} to exist, M must be a square matrix; hence for k parameters there must be $k + 1$ guesses for each parameter. Thus, both g and h must vary from 1 to k. It should also be noted that the partial derivatives in F_p are taken at some suitable point p between m_{0g} and m_{hg}, selected so that the Law of Mean Value holds. Let

y_i ($i = 1, \cdots, n$) designate the set of observed y values,

m_{bg} designate the parameter values which give the best fit,

Y_{bi} designate the corresponding values for the best fitting values of the Y's.

We may now define the following *row* vectors:

m with k elements $m_{bg} - m_{0g}$ designates the correction needed to change the first (or zero-th) guess into the best b guess,

c with n elements $Y_{bi} - Y_{0i}$ designates the corresponding changes in the calculated Y_0's,

d with n elements $y_i - Y_{0i}$ designates the difference between the zero-th approximation and the observed values, and

e with n elements $y_i - Y_{bi}$ designates the error of fit for the best values.

The problem may now be stated as follows: solve for the vector m in terms of Y, M, and d so that ee' is a minimum.

From the definition of elements we see that

$$e = d - c; \tag{25}$$

also

$$c = mF_q\,, \tag{26}$$

where q designates a suitable point on the curve between m_{0q} and n_{bq}. Thus,

$$e = d - mF_q\,. \tag{27}$$

Selecting m so as to minimize ee' is the multiple regression problem, which is solved as indicated in (16, pp. 173–174). Following this procedure,

$$ee' = dd' - dF_q'm' - mF_q\,d' + mF_qF_q'm'. \tag{28}$$

Differentiating with respect to m and setting the result equal to zero gives

$$2mF_qF_q' - 2dF_q' = 0. \tag{29}$$

If the changes in parameter values are slight so that F_p is approximately equal to F_q, then substituting $M^{-1}Y$ for F_q and solving for m gives

$$m = dY'(YY')^{-1}M. \tag{30}$$

Thus, we have an approximation for the correction term m expressed in terms of known values of the trial parameters M, Y, and d, the observed and predicted values of the dependent variables.

From the correction term m we can obtain a new vector of trial parameters $m_{(k+1)g}$, which should give a set of predictions $Y_{(k+1)i}$ which is better than any of the predictions Y_{oi} to Y_{ki} previously obtained. The new vector $m_{(k+1)g}$ can be substituted in the matrix M for the vector m_{kg} giving the poorest fit, and the resulting set of $k + 1$ trial parameters used to obtain a second m vector by the use of (30).

This solution exhibits the critical requirements of the Method of False Position much more clearly than does the lengthier expression in terms of elementary algebra given by Pearson (10). Since YY' must have an inverse, the rank must be k; that is, the different trial values of the parameters must be such that the result from one trial is not a linear function of the results from other trials. Correspondingly, since $Y = MF$, the various trial values of the parameters must be independent of each other, since YY' will have a rank less than k if the rank of M is less than k.

Thus, to the requirements stated at the beginning of this section, we must now add that the k changes in trial parameters must be linearly independent of each other and must result in a set of linearly independent Y's. For example, changing only one parameter for each set would result in a diagonal matrix for M which would clearly have a rank of k.

The procedure is an iterative one. It may be necessary to apply it several times to find a minimum. It also is desirable to check in the vicinity of the minimum to be sure that the point found is approximately a minimum. Since the function being minimized is a quadratic, it will have only one minimum.

Furthermore, the parameter changes and changes in error must be small

enough so that a line, plane, or hyperplane is a good fit to the surface in the region being dealt with. Under these conditions the equation $Y = MF$ will be a good approximation to the surface. Changing only one parameter at a time, so that M is a diagonal matrix, would be a simple method of satisfying this requirement.

The Method of False Position with some of its extensions and limitations has been discussed by Willers (17). He also presents other methods for solving problems of the type considered here.

Predicting Scale Values of Composite Stimuli

The four laws (equations 19, 20, 21, and 24) were used to predict the scale values of the composite stimuli from the scale values of their component stimuli.

For the negative exponential law four sets of values were chosen for B^-, B^+, and C. These values were used with (24) to give sets of values for $\hat{s}_{ij}$. The parameter values were used to construct the matrix M. The values of $\hat{s}_{ij}$ gave matrix Y. Using s_{ij} and $\hat{s}_{ij}$ gave the vector d. Equation (30) is then used to find a correction which gives a fifth set of values of the parameters, which is better than any of the first four. This process is repeated until a minimum is found and tested. For a one-parameter system the process is similar but much simpler. To give a measure of goodness of fit we have presented the sum of the squares of the discrepancies as well as the sum of the absolute values of the discrepancies. These values are shown in Table 3.

It can be seen that the logarithmic and square-root laws in this case give the largest discrepancies between the actual values of the composites and the values as predicted from the single stimuli. Therefore, we shall not consider either the logarithmic or square-root laws in further detail.

Both the linear and negative exponential laws placed the zero point

TABLE 3

Four Value Laws Compared

Law	Asymptote Pos. B^+ or $A^+ + C$	Asymptote Neg. B^- or $A^- + C$	Additive Constant C	$\Sigma\lvert s_{ij} - \hat{s}_{ij}\rvert$	$\Sigma(s_{ij} - \hat{s}_{ij})^2$
Negative Exponential	4.6	-1.8	+0.8	1.069	.205
Linear	∞	-∞	+0.94	1.658	.434
Logarithmic	∞	-∞	+1.10	2.177	.568
Square Root	∞	-∞	+1.13	2.487	.815

between pork and lamb, where previous consideration showed it reasonably came.

Figure 1 shows the test for the linear value increase—the plot of the

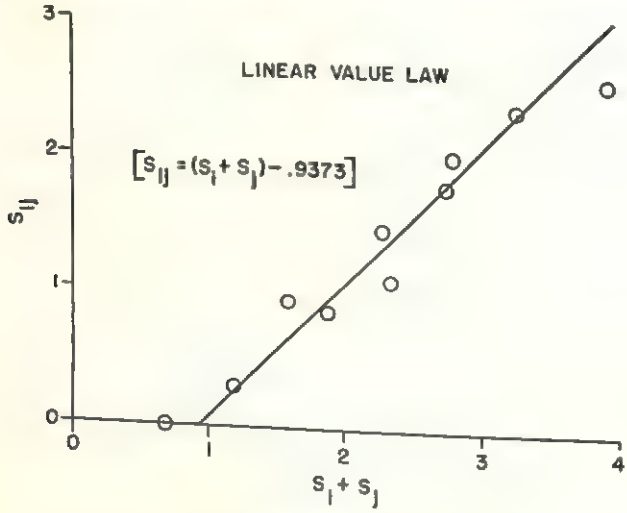


FIGURE 1

scale value of the composite against the sum of the scale values of the components. This plot gives a reasonably good fit to the line

$$\hat{s}_{ij} = (s_i + s_j) - .9373.$$

Thus, the best estimate of the zero point on the assumption of the linear law is .9373, or about .94 if we assume that $x_0 = 0$.

A more detailed analysis showing the fit for the negative exponential is shown in Figure 2. Here we have the value of tongue, .137, plotted on the abscissa, and over it the value of each of the composites with tongue—tongue and pork, tongue and lamb, tongue and beef, and tongue and steak. The same has been done for pork, lamb, beef, and steak and their composites with each of the other four stimuli.

The lines show how a negative exponential rule would fit the data, given that the zero point is at .8, that the upper asymptote is at 4.6 and the lower one at -1.8. We have a family of five lines, the upper line indicating the values v_{ij} for all composites with steak, designated S. The lower line indicates the values v_{ij} for all composites with tongue, designated T. Correspondingly, the other three lines show the values for composites with pork, lamb, and beef. These lines all converge at the two points (4.6, 4.6) and $(-1.8, -1.8)$, corresponding to the two asymptotes. It can be seen that the fit is good.

These data are not adequate to discriminate between the different

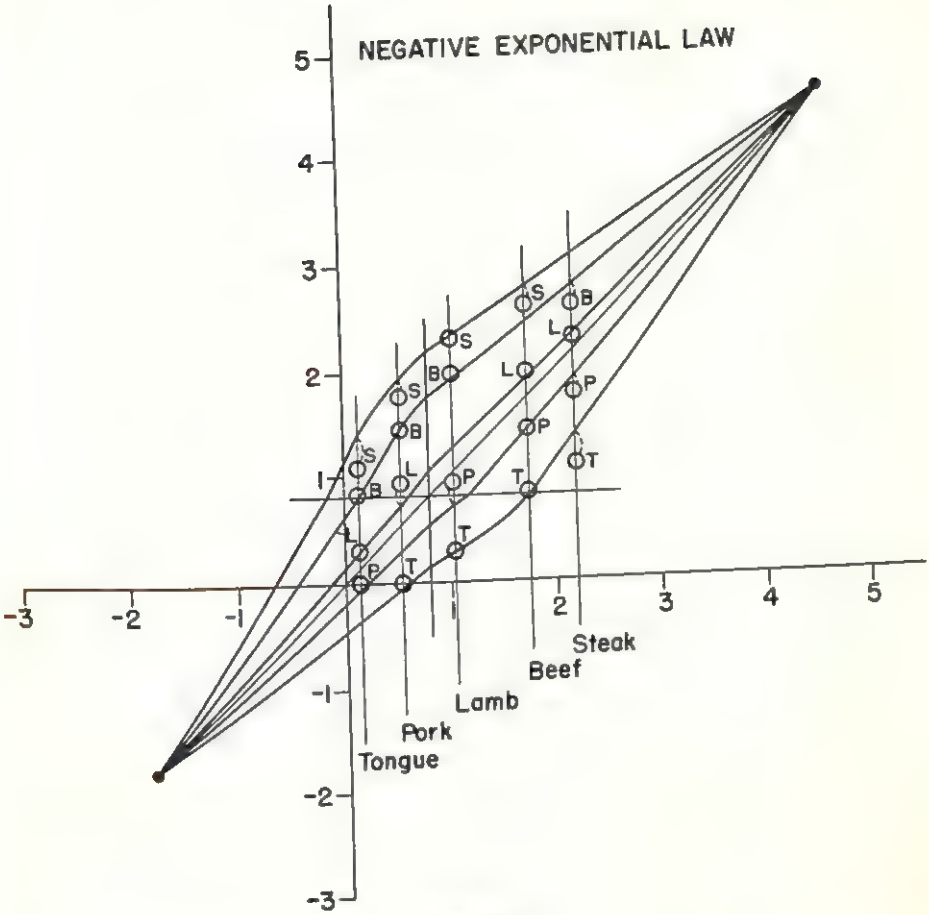


FIGURE 2

value laws even though the discrepancies are smallest for the negative exponential. More points are needed, particularly more points in the upper right quadrant, e.g., where both components are positive and hence the composite is positive. Similarly, if one has negative values, there should be more of them, so that the negative components would combine to form a number of different negative composites. It had been expected that all the values in this case would be positive and hence that five components might have been adequate.

Summary

Four different value laws have been developed: a square-root, a logarithmic, a negative exponential, and an additive law. A method has been presented for testing each law using only the scale values of components and composite stimuli determined by a psychophysical scaling method. In this case paired comparisons was used. A tentative extension of these rules has

been made to cover the case in which both negative and positive components are combined.

It has been shown that either the linear or the negative exponential law gives a good fit to limited data on food preferences. Also, it should be noted that persons do make consistent judgments about preferences for composite stimuli of the type used here, so that it is experimentally feasible to secure consistent judgments involving (i and j vs. g) or (i and j vs. g and h) in addition to the usual (i vs. j) type of choice.

Thus, we have a procedure for investigating the laws governing preferences, or value judgments, in areas where there is no readily available method for obtaining a physical measure of the amount of the commodity, and in which the usual scaling methods do not give a zero point.

REFERENCES

1. Bishop, Ruth. Points of neutrality in social attitudes of delinquents and non-delinquents. *Psychometrika*, 1940, **5**, 35-45.
2. Cauchy, Augustin Louis. Analyse algébrique, 1821. (Chapter V, Problem I.)
3. Gulliksen, H. A rational equation of the learning curve based on Thorndike's law of effect. *J. gen. Psychol.*, 1934, **11**, 395-434.
4. Gulliksen, H. Paired comparisons and the logic of measurement. *Psychol. Rev.*, 1946, **53**, 199-213.
5. Gulliksen, H. A least squares solution for paired comparisons with incomplete data. *Psychometrika*, 1956, **21**, 125-134.
6. Hamel, Georg. Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung $f(x + y) = f(x) + f(y)$. *Mathematische Annalen*, 1905, **60**, 459-462.
7. Horst, A. P. A method for determining the absolute value of a series of stimulus situations. *J. educ. Psychol.*, 1932, **23**, 418-440.
8. Mitscherlich, E. A. Das Gesetz des Minimums und das Gesetz des abnehmenden Bodenertrages. *Landwirtschaftl. Jarhb. Zsch. f. wiss. Landwirtschft*, 1909, **38**, 537-552.
9. Osgood, W. F. Advanced calculus. New York: MacMillan, 1940.
10. Pearson, Karl. On a general theory of the method of false position. *Phil. Mag.*, 1903, **4**, 658-668 (6th series).
11. Sierpiński, Waclaw. Sur l'équation fonctionelle $f(x + y) = f(x) + f(y)$. *Fundamenta Mathematicae*, 1920, **1**, 116-122.
12. Spillman, W. J. and Lang, E. The law of diminishing returns. Yonkers, N. Y.: World Book Co., 1924.
13. Thurstone, L. L. The indifference function. *J. soc. Psychol.*, 1931, **2**, 139-167.
14. Thurstone, L. L. Measurement of values. *Psychol. Rev.*, 1954, **61**, 47-58.
15. Thurstone, L. L. and Jones, L. V. The rational origin for measuring subjective values. (Unpublished manuscript.)
16. Turnbull, H. W. and Aitken, A. C. An introduction to the theory of canonical matrices. London and Glasgow: Blackie and Son, Limited, 1950.
17. Willers, F. A. (Trans. by R. T. Beyer) Practical analysis graphical and numerical methods. New York: Dover Publ. Inc., 1948.

Manuscript received 4/7/55

Revised manuscript received 10/3/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

MAXIMIZING TEST BATTERY PREDICTION WHEN THE WEIGHTS ARE REQUIRED TO BE NON-NEGATIVE

JOSEPH LEV

NEW YORK STATE DEPARTMENT OF CIVIL SERVICE

A procedure is developed for computing optimum regression weights under the restriction that they be non-negative. The weights maximize, subject to the restrictions, the multiple correlation between several predictors and a criterion. A numerical example is provided.

For some purposes weights of subtests of an examination must be positive. This is true, for example, in civil service examinations when weights are announced in advance of administration of the examinations. Positive weights are needed in order to motivate candidates to perform well in the subtests. The problem considered in this paper is the computation of weights of subtests so as to maximize prediction of a criterion, with the restriction that the weights be non-negative.

The procedure to be described is similar to the method of "steepest descent" (e.g., 1, p. 47). Actually it is more appropriate to call this a method of steepest ascent, since the weights are determined by successive addition of increments which tend to increase the multiple correlation between the predictors and the criterion.

An iterative procedure to accomplish the purpose considered in this paper was previously described in (2). Advantages of the present method are that the computations at each step in the iteration indicate the procedure at the next step, the termination of the procedure is clearly indicated, and the weights obtained are the best possible.

In order to write the necessary formulas, the following notation will be adopted:

X_0 = criterion
X_i = predictors ($i = 1, 2, \cdots, n$)
a_i = weights for predictors ($i = 1, 2, \cdots, n$)
$T = a_1X_1 + a_2X_2 + \cdots + a_nX_n$ = total weighted predictor score
$C_{ii} = \sum (X_i - \bar{X}_i)^2$ = sum of squares for X_i
$C_{ij} = \sum (X_i - \bar{X}_i)(X_j - \bar{X}_j)$ = sum of products for X_i and X_j
R = correlation between X_0 and T

The L_{ij} used by Wherry and Gaylord may be used here in place of the C_{ij}. The C_{ij} are smaller numbers than the L_{ij} and are sufficient for the

accuracy required. In this notation

$$R = \frac{\sum_{1}^{n} a_i C_{0i}}{\sqrt{C_{00}} \sqrt{\sum_{i,j=1}^{n} a_i a_j C_{ij}}}. \tag{1}$$

In (1) R may be viewed as a function of the variables a_i ; the C_{ij} are constants. It is convenient to introduce the total differential

$$dR = \frac{\partial R}{\partial a_1} da_1 + \frac{\partial R}{\partial a_2} da_2 + \cdots + \frac{\partial R}{\partial a_n} da_n . \tag{2}$$

Here dR is an increment in R, da_i are increments in the weights, and $\partial R/\partial a_i$ are partial derivatives. (2) shows how changes in weights influence changes in R. If partial derivatives have been computed for fixed values of the a_i , values of the da_i can be selected so as to make dR positive; that is, increments can be chosen for the weights so as to cause an increase in R. If increments are chosen corresponding to large values of the partial derivatives, the increase in R to the maximum can be made with few iterations. For this reason the present procedure is a method of "steepest ascent."

For computation of the partial derivatives we note that

$$\frac{\partial R}{\partial a_i} = (C_{00} \sum_{i,j} a_i a_j C_{ij})^{-1/2} [C_{0i} - k \sum a_j C_{0j}], \tag{3}$$

where $k = \sum a_j C_{0j} / \sum_{i,j} a_i a_j C_{ij}$.

In the computations to be presented, only relative values of the partial derivatives are required. Consequently, only the quantities within the square brackets are computed.

In essence the computation involves repeated trials of increments in the weights, da_i , so that each trial produces a positive increment, dR, in R with the restriction that the a_i be positive. Values of the da, selected at any trial are determined by the values of the partial derivatives computed at the previous trial. The trials are terminated when all of the partial derivatives are either zero (or nearly zero) or definitely negative. At this stage the terminal weights a_{ti} are computed as the sums of the increments tried at the various iterations. For those variables which have zero partial derivatives the terminal weights should be positive or zero. For the variables which have negative partial derivatives, the terminal weights should be zero.

It is evident that a maximum in R has been reached at the termination point described above. For the variables having zero partial derivatives, the terminal values of the a_i provide the usual maximum in R. For the remaining variables an increase in R can be obtained only by use of negative weights.

A justification for the procedure will be discussed more fully after the computations have been described.

Computational Procedures

The computing procedures will be described first symbolically, and then in relation to a numerical application. The symbolic development which follows is outlined in Table 1.

1. Record the sums of squares and cross products of the predictor and criterion variables as a square matrix.

2. It is helpful, though not essential, to record below this matrix the correlations between the predictors and the criterion.

3. Below the matrix record the increments corresponding to the iterations in the appropriate columns. All the increments corresponding to one iteration are placed in a single row. The symbol da_{ij} will be used to denote the increment for the variable X_j in the ith iteration.

4. After m iterations the weight a_{mj} for variable X_j is the sum of the increments used to this point,

$$a_{mj} = \sum_{i=1}^{m} da_{ij} .$$

The terminal weight for X_j is the a_{tj} after all iterations have been completed.

5. The first increment is obtained for the variable, X_j , which has highest correlation with the criterion. It is an approximation of one or two digits to the ratio C_{0j}/C_{jj} . Increments for the remaining variables are ordinarily, though not necessarily, taken to be zero in the first iteration. These increments are also the weights after the first iteration. That is, $a_{1j} = da_{1j}$ for each predictor X_j .

6. The computations which lead to the next iteration are carried out in columns set up at the right of the matrix of sums of squares and cross products.

7. In the first column of this set-up the following computations are made successively:

$$\sum a_{1j}C_{1j} , \sum a_{1j}C_{2j} , \cdots , \sum a_{1j}C_{nj} , \sum a_{1j}C_{0j} ,$$

$$\sum_i \sum_j a_{1i}a_{1j}C_{ij} = \sum_i a_{1i}(\sum_j a_{1j}C_{ij}) \qquad (i, j = 1, 2, \cdots , n),$$

$$k_1 = (\sum a_{1j}C_{0j})/(\sum \sum a_{1i}a_{1j}C_{ij}),$$

$$C_{00}R^2 = k_1 \sum a_{1j}C_{0j} . \tag{4}$$

Formula (4) provides a measure of the correlation between the weighted sum and the criterion, attained as a result of the first iteration. A similar quantity is computed at each iteration.

TABLE 1

SYMBOLIC LAY-OUT FOR COMPUTATION OF WEIGHTS, SHOWING TWO ITERATIONS

	X_1	. . . .	X_n	X_0	I	II
X_1	C_{11}	. . .	C_{1n}	C_{01}	$\Sigma\, a_{1j}C_{1j}$	$\Sigma\, a_{2j}C_{1j}$
.						
.						
.						
X_n	C_{n1}		C_{nn}	C_{0n}	$\Sigma\, a_{1j}C_{nj}$	$\Sigma\, a_{2j}C_{nj}$
X_0	C_{01}		C_{0n}	C_{00}	$\Sigma\, a_{1j}C_{0j}$	$\Sigma\, a_{2j}C_{0j}$
	r_{01}		r_{0n}		$\Sigma\Sigma\, a_{1i}a_{1j}C_{ij}$	$\Sigma\Sigma\, a_{2i}a_{2j}C_{ij}$
I	da_{11}		da_{1n}		$k_1 = \dfrac{\Sigma\, a_{1j}C_{0j}}{\Sigma\Sigma\, a_{1i}a_{1j}C_{ij}}$	$k_2 = \dfrac{\Sigma\, a_{2j}C_{0j}}{\Sigma\Sigma\, a_{2i}a_{2j}C_{ij}}$
II	da_{21}		da_{2n}			
.						
.					$k_1\, \Sigma\, a_{1j}C_{0j}$	$k_2\, \Sigma\, a_{2j}C_{0j}$
.						
t	da_{t1}		da_{tn}		$C_{01} - k_1\, \Sigma\, a_{1j}C_{1j}$	$C_{01} - k_2\, \Sigma\, a_{2j}C_{1j}$
			.		.	
					.	
Total	a_{t1}		a_{tn}			
Regression weight	$k_t a_{t1}$		$k_t a_{tn}$		$C_{0n} - k_1\, \Sigma\, a_{1j}C_{nj}$	$C_{0n} - k_2\, \Sigma\, a_{2j}C_{nj}$

The succeeding computations have as their purpose an evaluation of the partial derivatives exhibited in (2). We compute

$$C_{01} - k_1 \sum a_{1j}C_{1j} , \cdots , C_{0n} - k_1 \sum a_{1j}C_{nj} .$$

For the first column some simplification of the above computations is possible. However, the saving is slight and the formulas are stated in full generality as they apply to all columns. Note that the partial derivative which corresponds to the variable having positive weight becomes zero, or nearly so.

8. The increment for the second iteration is based on the values of the partial derivatives obtained from the first iteration. Ordinarily we choose the partial derivative which has the highest positive value and select an increment for the corresponding variable. If this variable is X_h , then the increment da_{2h} is chosen so that

$$C_{0h} - \sum_j a_{2j}C_{hj} = 0$$

approximately. Here the a_{2j} are the weights after the second iteration. *Increments for more than one predictor may be chosen at any iteration if this seems appropriate.*

9. To perform the computations for the second iteration we take advantage of the relationship

$$\sum_j a_{2j}C_{ij} = \sum_j a_{1j}C_{ij} + \sum_j (da_{2j})C_{ij} .$$

However, this simplification is not appropriate for the computation of $\sum \sum a_{2i}a_{2j} C_{ij}$. The remaining computations are carried out as described for the first iteration.

10. The iterations are continued until the partial derivatives are decidedly negative, or approximately zero, and the quantity $k_m \sum a_{mj}C_{0j}$ seems to have attained a maximum.

Details of a numerical example are shown in Table 2. The first positive increment, $da_{12} = 0.6$, is an approximation to the quotient 1790/3246. This increment is also a_{12} . The computations based on this increment are shown in the column headed I. The first seven entries in this column are products of the increment .6 and the cross products under X_2 as .6 (2802) = 1681, $\cdots$, .6 (1790) = 1074. The eighth entry is .6 (1948) = 1169; k_1 is 1074/1169 = .9187; $k_1 \sum a_{1j}C_{0j} = C_{00}R^2$ is given as .9187 (1074) = 986.7. The last six entries in this column indicate the relative values of the partial derivatives. The greatest of these is the one related to X_6 and has the value 2520 − (.9187) (1654) = 1000. This value of the partial derivative suggests the positive increment $da_{26} = 0.1$ for the second iteration. Computations in the

TABLE 2

ITERATIVE COMPUTATION OF WEIGHTS

	X_1	X_2	X_3	X_4	X_5	X_6	X_0	Formula		I	II	III	IV	V
X_1	7124	2802	1046	1083	320	1794	1254	$\Sigma\, a_j C_{1j}$		1681	1860	1878	1887	1892
X_2	2802	3246	1281	1617	570	2757	1790	$\Sigma\, a_j C_{2j}$		1948	2224	2252	2266	2274
X_3	1046	1281	2025	1347	1713	1581	534	$\Sigma\, a_j C_{3j}$		769	927	943	951	958
X_4	1083	1617	1347	3752	2499	2206	979	$\Sigma\, a_j C_{4j}$		970	1191	1213	1224	1243
X_5	320	570	1713	2499	9518	-1263	-65	$\Sigma\, a_j C_{5j}$		342	216	203	197	209
X_5	1794	2757	1581	2206	-1263	13380	2520	$\Sigma\, a_j C_{6j}$		1654	2992	3126	3193	3204
X_0	1254	1790	534	979	-65	2520	12438	$\Sigma\, a_j C_{0j}$		1074	1326	1351	1364	1369
r	.1332	.2817	.1063	.1434	-.0060	.1954		$\Sigma\Sigma\, a_i a_j C_{ij}$		1169	1634	1695	1727	1739
								k		.9187	.8115	.7971	.7898	.7872
I	0	.6	0	0	0	0		$k\,\Sigma\, a_j C_{0j}$		986.7	1076.0	1076.9	1077.3	1077.7
II	0	0	0	0	0	.1		$C_{01} - k\,\Sigma\, a_j C_{1j}$		-291	-256	-243	-236	-235
III	0	0	0	0	0	.01		$C_{02} - k\,\Sigma\, a_j C_{2j}$		0	-16	-5	0	0
IV	0	0	0	0	0	.005		$C_{03} - k\,\Sigma\, a_j C_{3j}$		-173	-219	-218	-217	-220
V	0	0	0	.005	0	0		$C_{04} - k\,\Sigma\, a_j C_{4j}$		-88	12	12	12	1
Total	0	.6	0	.005	0	.115		$C_{05} - k\,\Sigma\, a_j C_{5j}$		-379	-240	-227	-221	-230
Regression weight	0	.4723	0	.0039	0	.0905		$C_{06} - k\,\Sigma\, a_j C_{6j}$		1000	90	28	-2	-2

column headed II will be given in full. Notice that

$$
\begin{aligned}
1681 + .1(1794) &= 1860 \\
1948 + .1(2757) &= 2224 \\
769 + .1(1581) &= 927 \\
970 + .1(2206) &= 1191 \\
342 + .1(-1263) &= 216 \\
1654 + .1(13380) &= 2992 \\
1074 + .1(2520) &= 1326 \\
.6(2224) + .1(2992) &= 1634 \\
1326/1634 &= .8115 \\
.8115\ (1326) &= 1076.0 \\
1254 - (.8115)\ (1860) &= -256 \\
1790 - (.8115)\ (2224) &= -16 \\
534 - (.8115)\ (927) &= -219 \\
979 - (.8115)\ (1191) &= 12 \\
-65 - (.8115)\ (216) &= -240 \\
2520 - (.8115)\ (2992) &= 90
\end{aligned}
$$

Iterations and computations continue as shown in Table 2. At iteration 5 the process stops because the partial derivatives for variables X_2, X_4, X_6 are nearly zero and the remaining partial derivatives are negative. Note also the stability of $C_{00}R^2$ over the last three iterations. The result is a multiple R of .2994, with regression weights 0, .4723, 0, .0039, 0, .0905.

All the increments tried in this numerical example have been positive. In some circumstances negative increments may be tried. This is true when a positive weight has been too large as shown by a negative partial derivative for a variable having positive weight.

An evaluation of the method is called for at this point. One may ask whether the weights obtained by this method actually provide the highest possible R under the restriction of non-negative weights. Before dealing with this question it is necessary to consider whether there exists a set of positive values a_i which yield a maximum for R.

One way of demonstrating that a maximum actually exists is to note first that under ordinary conditions a set of weights maximizing R can be found for any selection of variables out of the n given variables. Some of the selections of variables will give a maximum R with non-negative weights. There are a finite number of these selections. Consequently, we may choose that one set of variables which provides the largest maximum R under the required condition. Weights obtained in this way are the desired weights. Thus, the existence of a maximum is demonstrated. In exceptional circumstances it may happen that all weights are negative. This will be shown by the computation.

The existence of a maximum R under the condition of non-negative

weights having been established, it is necessary to demonstrate that the procedure described in this paper actually leads to this maximum. This will be demonstrated with the aid of a theorem.

The necessary and sufficient condition for the set of weights $a_1, a_2, \cdots, a_n$ to maximize R, with the provision that none of the weights be negative, is that the partial derivatives $\partial R/\partial a_i$ be zero for variables X_i with positive weights, and negative or zero for variables X_i with zero weights.

Necessity: This has already been discussed. If a set of non-negative values of the a_i maximizing R has been attained then the condition implies that R can be further increased only by changing from zero to negative some of the a_i, for variables having negative partial derivatives. Any change in values of the a_i for variables which have zero partial derivatives will decrease R. Thus necessity is demonstrated.

Sufficiency: To demonstrate sufficiency, suppose that the procedure of this paper has provided a set of a_i satisfying the conditions of the theorem with corresponding R. Suppose now that there is another set of values of the a_i also satisfying the conditions of the theorem with a value of $R' > R$. But this is impossible, for some of the partial derivatives for the first set of a_i should then be positive. Thus sufficiently is also demonstrated.

Since the procedure described in this paper leads to zero or negative partial derivatives, the procedure provides a unique maximum under the restriction to non-negative weights.

REFERENCES

1. Householder, A. S., Principles of numerical analysis. New York: McGraw-Hill, 1953.
2. Wherry, R. J. and Gaylord, R. H. Test selection with integral gross score weights. *Psychometrika*, 1946, 11, 173-183.

Manuscript received 5/27/55

Revised manuscript received 10/17/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

THE MATCHING PROBLEM*

EDGAR J. GILBERT

HARVARD UNIVERSITY†

Tables of the exact distributions of number of matches are given for small decks having the same number of cards in each suit. Several approximate distributions are considered for use with larger decks, and some indication of the goodness of the approximations is given.

Many psychological experiments involve testing the ability of a person (or a method) to classify certain objects into categories. The matching-problem technique to be described in this paper provides a mathematical model which may be used to calculate significance levels for such experiments.

The example of the matching problem which is most often cited is the work done in certain of the early ESP (extra-sensory perception) experiments. A deck of 25 cards containing 5 each of 5 different figures (circle, cross, wave, square, and star) was shuffled into a random order by the experimenter. The subject was given a second deck of the same composition and asked to arrange it in the same order as the hidden deck of the experimenter. Then the two decks were compared. If the first card of the subject's deck was of the same kind (circle, cross, etc.) as the first card of the experimenter's deck, the subject scored a "match." Then the second card in each deck was examined, and so on through the two decks. The subject's ability was scored according to the total number of correct matches, as compared to the number to be expected by chance alone.

For another example, suppose a handwriting expert claims he can tell a person's profession by examining a sample of his handwriting. To test his ability he is given 10 samples, 2 written by doctors, 4 by lawyers, and 4 by teachers. He is told which professions are represented but not how many samples are from each profession. If we number the samples from 1 to 10, placing under each number the true profession and then the expert's guess, we might get the following result:

Sample Number:	1	2	3	4	5	6	7	8	9	10
True Profession:	L	D	L	T	D	T	L	T	T	L
Expert's Guess:	L	D	T	T	D	T	D	L	D	D

*The calculations for this paper were done while the writer was working on a project sponsored by funds from the Ford Foundation. The writer wishes to express his gratitude to Dr. Frederick Mosteller for his constant help and encouragement during the writing of this paper.

†Now at University of California, Berkeley.

In order to evaluate the expert's ability on the basis of this test, we might think of the two rows "True Profession" and "Expert's Guess" as being the arrangements of a target deck and a call deck, respectively, and imagine that the expert was attempting a kind of matching problem similar to that of the ESP experiment described above. We notice that he called correctly cards numbered 1, 2, 4, 5, and 6, for a total of 5 correct matches. However, we see that this case is different from the ESP case in that the deck used by the expert does not have the same composition as that used by the experimenter; so we must allow for this in calculating the probability of his getting the 5 matches by chance. The probabilities for this particular example will be worked out in the last section of the paper, but first the more easily calculated cases where both decks are identical will be considered.

In general, the model will consist of two decks of cards: a target deck, the suits of which correspond to the kinds of objects being classified, and a call deck whose suits correspond to the categories used by the subject in his classification. We imagine that the cards in the target deck have some arbitrary arrangement and calculate the probability distribution of the number of correct matches between cards in the two decks on the assumption that the call deck was arranged by random shuffling. If the subject has any ability, scores as high as or higher than his will have low probability.

Several variations of the problem arise, depending on the composition of the call deck. If the person to be tested is informed of the categories to be used and the number of objects in each category, and is forced to use this information, then the call deck is identical with the target deck in composition. If he is not so informed or not forced to use the information, he may classify the objects in such a way that his call deck is different from the target deck. In this case, his ability to choose categories is not perfect. In order to use the matching problem technique, the calculation of the probability of various numbers of matches must be based on the composition of the call deck actually used. It is always possible to assume that both decks have the same suits, however, by using the trick of saying a suit which is not represented actually is present, but with zero cards in it.

In this paper, tables of the exact distributions of numbers of matches for small decks having the same number of cards in each suit will be given. Several methods of obtaining approximate distributions for larger decks will be indicated. The problems of decks with different numbers of cards in different suits, and of non-identical decks, will be discussed.

Exact Distributions

The simplest form of the matching problem is the case where both the target deck and the call deck have n distinct cards, i.e., where any given card of the call deck can match one and only one card in the target deck. This is a special form of a problem which has a long history in the mathe-

matical literature. For example, Feller (**3**) gives a formula for calculating the probability distribution of the number of matches in this case and gives tables of the calculated values for decks of size 3, 4, 5, 6, and 10. He shows that for deck sizes larger than 10, the values are, to five decimal places, the same as for the Poisson distribution with unit mean.

A special case of the matching problem arises when there are only two suits in each deck and each suit has c cards in it, making a total of $2c$ cards in the deck. There is no possible arrangement of the call deck which will give an odd number of matches, while the number of arrangements giving h matches, when h is an even integer, is $\binom{c}{h/2}^2$, the square of the number of combinations of c things taken $h/2$ at a time. Since this is the square of a binomial coefficient, it is a simple matter to look up the binomial coefficients corresponding to $h = 0, 2, 4, \cdots 2c$, square them, and add the results to get the total number of ways of arranging the deck. Then divide the number of arrangements which give h matches by the total number of arrangements to get the probability of exactly h matches. For example, if each suit has 2 cards in it, then there are $\binom{2}{0}^2 = 1$ way of getting 0 matches, $\binom{2}{1}^2 = 4$ ways of getting 2 matches, and $\binom{2}{2}^2 = 1$ way of getting 4 matches. So the probability of exactly h matches when $h = 0, 1, 2, 3, 4$ is 1/6, 0, 4/6, 0, 1/6.

The next easiest case to calculate is that of two identical decks, each having s suits of c cards per suit. It is this situation, where the probability of exactly h matches is a function of the three numbers: s, the number of suits; c, the number of cards in each suit; and h, the number of matches, which will be the primary concern in this paper. Denote this probability by $m(s, c; h)$ and write the probability of h or more matches (which is the sum of the probabilities for h, $h + 1$, $h + 2$, etc.) as $M(s, c; h)$. It is this latter probability which is usually wanted in finding the significance of experimental results. We usually want to know the probability of a subject's doing as well as or better than his result by chance alone.

Even in this case, few tables have been published giving the exact probabilities. Huntington (**8**) gives tables of $m(3, 3; h)$ and $m(4, 4; h)$. Greville (**5**) gives a table of $m(5, 5; h)$, the probabilities for the ESP experiment referred to earlier. Greenwood (**4**) gives a table of estimated values of $m(4, 13; h)$ for values of h from 0 to 7, where both decks are the ordinary 52-card bridge decks. There may be other tables available in the literature; but it was thought that a small collection of tables in one place might be useful.

Greville (**6**) derived a formula for the matching problem distribution which seemed to the writer to be more adapted to calculation of exact probabilities than some others in the literature. His formula was for a more general case than that we are now considering and will be used in its general form in

TABLE 1

Exact Distributions. Probability of h or More Matches

h	M(2, 1;h)	M(3, 1;h)	M(4, 1;h)	M(5, 1;h)	M(6, 1;h)	M(7, 1;h)	M(8, 1;h)	M(9, 1;h)	M(10, 1;h)	M(11, 1;h)
0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1	.50000	.66667	.62500	.63333	.63194	.63214	.63212	.63212	.63212	.63212
2	.50000	.16667	.29167	.25833	.26528	.26409	.26426	.26424	.26424	.26424
3		.16667	.04167	.09167	.07778	.08075	.08023	.08031	.08030	.08030
4			.04167	.00833	.02222	.01825	.01912	.01897	.01899	.01899
5				.00833	.00139	.00436	.00350	.00369	.00366	.00365
6					.00139	.00020	.00072	.00056	.00060	.00059
7						.00020	.00002	.00010	.00008	.00008
8							.00002	.00000	.00001	.00001

h	M(2, 2;h)	M(3, 2;h)	M(4, 2;h)	M(5, 2;h)	M(6, 2;h)	M(7, 2;h)	M(8, 2;h)	M(9, 2;h)	M(10, 2;h)	M(11, 2;h)
0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1	.83333	.88889	.88214	.87870	.87632	.87462	.87335	.87237	.87158	.87094
2	.83333	.62222	.61548	.60991	.60689	.60483	.60334	.60220	.60131	.60060
3	.16667	.32222	.32341	.32451	.32406	.32383	.32369	.32360	.32354	.32350
4	.16667	.14444	.13294	.13262	.13450	.13580	.13676	.13749	.13806	.13853
5		.01111	.03532	.04179	.04396	.04538	.04640	.04716	.04776	.04824
6		.01111	.00992	.01033	.01144	.01226	.01285	.01330	.01365	.01394
7			.00040	.00177	.00237	.00270	.00295	.00313	.00328	.00340
8			.00040	.00036	.00041	.00049	.00056	.00062	.00067	.00071
9				.00001	.00005	.00007	.00009	.00010	.00012	.00013
10				.00001	.00001	.00001	.00001	.00002	.00002	.00002

h	M(2, 3;h)	M(3, 3;h)	M(4, 3;h)	M(5, 3;h)	M(6, 3;h)	M(7, 3;h)	M(8, 3;h)	M(2, 4;h)	M(3, 4;h)	M(4, 4,h)
0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1	.95000	.96667	.96258	.96016	.95852	.95734	.95645	.98571	.99001	.98813
2	.95000	.83810	.82875	.82266	.81875	.81603	.81402	.98571	.93737	.93005
3	.50000	.61310	.60010	.59456	.59113	.58882	.58715	.75714	.80646	.79355
4	.50000	.35417	.35404	.35335	.35313	.35301	.35295	.75714	.60052	.58930
5	.05000	.16131	.16724	.17154	.17407	.17578	.17702	.24286	.37117	.37156
6	.05000	.04881	.06322	.06808	.07113	.07319	.07467	.24286	.18831	.19645
7		.01667	.01864	.02215	.02425	.02569	.02674	.01428	.07472	.08658
8		.00060	.00461	.00592	.00694	.00766	.00820	.01428	.02485	.03174
9		.00060	.00073	.00130	.00167	.00195	.00217		.00511	.00962
10			.00015	.00023	.00034	.00043	.00050		.00141	.00242
11			.00000	.00004	.00006	.00008	.00010		.00003	.00049
12			.00000	.00000	.00001	.00001	.00002		.00003	.00009
13										.00001

TABLE 1 (Continued)

h	M(5,4;h)	M(2,5;h)	M(3,5;h)	M(4,5;h)	M(5,5;h)	M(2,6;h)	M(3,6;h)	M(2,7;h)	M(3,7;h)	M(2,8;h)
0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1	.98693	.99603	.99702	.99624	.99571	.99892	.99911	.99971	.99974	.99992
2	.92558	.99603	.97700	.97286	.97026	.99892	.99186	.99971	.99720	.99992
3	.78646	.89682	.91278	.90271	.89690	.95996	.96331	.98543	.98532	.99495
4	.58385	.89682	.78150	.76776	.76033	.95996	.89136	.98543	.94969	.99495
5	.37138	.50000	.59102	.58113	.57600	.71645	.76155	.85693	.87266	.93403
6	.20090	.50000	.38388	.38408	.38406	.71645	.58399	.85693	.74532	.93403
7	.09232	.10397	.21077	.21933	.22380	.28355	.39331	.50000	.57839	.69036
8	.03609	.10397	.09571	.10764	.11374	.28355	.22923	.50000	.40070	.69036
9	.01202	.00397	.03624	.04528	.05043	.04004	.11416	.14306	.24472	.30964
10	.00342	.00397	.01074	.01630	.01953	.04004	.04838	.14306	.13051	.30964
11	.00083		.00281	.00500	.00662	.00108	.01709	.01457	.06040	.06597
12	.00017		.00043	.00131	.00196	.00108	.00513	.01457	.02404	.06597
13	.00003		.00010	.00029	.00051		.00120	.00029	.00823	.00505
14			.00000	.00005	.00012		.00026	.00029	.00237	.00505
15				.00001	.00002		.00003		.00059	.00008
16							.00001		.00011	.00008
17									.00002	

h	M(2,9;h)	M(2,10;h)	M(2,11;h)	M(2,12;h)
0	1.00000	1.00000	1.00000	1.00000
1	.99998	.99999	1.00000	1.00000
2	.99998	.99999	1.00000	1.00000
3	.99831	.99945	.99983	.99995
4	.99831	.99945	.99983	.99995
5	.97166	.98849	.99554	.99834
6	.97166	.98849	.99554	.99834
7	.82653	.91055	.95694	.98044
8	.82653	.91055	.95694	.98044
9	.50000	.67186	.80257	.88983
10	.50000	.67186	.80257	.88983
11	.17347	.32814	.50000	.65786
12	.17347	.32814	.50000	.65786
13	.02834	.08945	.19743	.34214
14	.02834	.08945	.19743	.34214
15	.00169	.01151	.04305	.11017
16	.00169	.01151	.04305	.11017
17	.00002	.00055	.00446	.01956
18	.00002	.00055	.00446	.01956

the last section of the paper. The following special case of Greville's formula was used to calculate the values given in Table 1: ($n = sc$ = size of deck)

$$m(s, c; h) = \frac{1}{n!} \sum_{i=h}^{n} (-1)^{i-h} \binom{i}{h} (n - i)! H_i , \tag{1}$$

where H_i is the coefficient of x^i in the expansion of

$$\left(\sum_{j=0}^{c} \frac{c!c!x^j}{j!(c-j)!(c-j)!} \right)^s . \tag{2}$$

Since it is usually the probability of h or more matches which is wanted in checking experimental data, Table 1 lists $M(s, c; h) = \sum_{i=h}^{n} m(s, c; i)$. The calculations were carried out to exact integers before dividing by $n!$, the divisions made to 6 decimals, and the values of $M(s, c; h)$ rounded to five decimals after summing. The table is arranged first by groups in order of increasing c, then within groups in order of increasing s. Within each group it will be noticed that the distributions seem to approach a limiting distribution—the Poisson—but that the approach is slower for groups with larger c. It will also be noticed that the first distribution in each group [that for $M(2, c; h)$] has the entry for odd h equal to the entry for the following even h. This is due to the fact mentioned earlier, that the probability in this case of exactly h matches, when h is odd, is zero.

The values $M(5, 5; h)$ given in Table 1 were obtained by summing to five decimals the values given to 7 decimals by Greville (**5**). All other values in Table 1 were calculated independently by the writer, using the tables of Feller and Huntington as a check where there were duplications.

Approximate Distributions

Since the calculation of exact probabilities gets very tedious for large decks, it would be desirable to be able to use distributions which either are already available, or are easier to calculate, to approximate significance levels for the number of matches. As has already been mentioned, Feller's tables (**3**) show that a Poisson distribution is satisfactory for the approximation of $m(s, 1; h)$, when s is only moderately large. A study of the distributions within each group with the same value of c (Table 1) leads one to suspect that this is true also of groups other than $c = 1$; but as c gets large, the rate of approach to Poisson with increasing s is much slower.

In order to find more accurate approximations, we need to know more about the moments of the matching problem distribution. We shall continue throughout this section to work with the relatively simple case of identical decks with the same number of cards in each suit. For this case, the first four moments were derived by means of a rather ingenious use of determinants by Olds (**9**). If ν is the mean and μ_i the ith moment about the mean,

then in terms of c, the number of cards per suit, and $n = sc$, the size of the deck, we have

$$\begin{aligned} \nu &= c \\ \mu_2 &= c(n - c)/(n - 1) \\ \mu_3 &= c(n - c)(n - 2c)/(n - 1)(n - 2) \\ \mu_4 &= \frac{c(n - c)}{(n - 1)(n - 2)(n - 3)} [(n - 2c)(n - 3c)(3c + 1) \\ &\qquad + (c - 1)(12nc - n - 18c^2 - 6c)]. \end{aligned} \tag{3}$$

Anderson (1) shows that as the deck size is increased while keeping the proportion of cards in each suit fixed, the number of matches is asymptotically normally distributed. So if n is large enough, the mean and variance of the matching problem distribution can be calculated from formulas (3), and the desired values found in a table of the normal distribution. Unfortunately, it is difficult to say precisely when n is large enough. In many cases, the matching problem distribution is asymmetrical; to the extent that this is so, the normal will give a bad fit. Although the normal is not as close an approximation as some of the distributions considered later in this section, it can be seen from the examples in Table 2 that it is sufficiently close for some purposes; it has the advantage of being available without too much calculation. An example of the calculations involved in fitting a normal to a matching problem is worked out in the last section of this paper.

Hamilton (7) noted that the matching problem distribution was somewhat like a binomial distribution with $p = 1/s$ and $n = sc$, and suggested that writing the mean of the matching problem as np and variance as $npq[n/(n - 1)]$ made the similarity apparent. It can be seen that the binomial with $p = 1/s$ has the same mean as the matching problem, and that the ratio of the two variances approaches 1 as n gets large. This is also the case with the third moment about the mean. It is $npq(2q - 1)$ for the binomial, and we can rearrange the formula for μ_3 in (3) above—substituting np for c where it appears and setting $q = 1 - p$—to make the third moment of the matching problem look like $npq(2q - 1)n^2/(n - 1)(n - 2)$.

The binomial is suggested as a good approximating distribution for another reason—extensive tables of the binomial are available (11). Since the binomial is a two-parameter distribution, we can in many cases find a tabled distribution which has both mean and variance equal to that of the matching problem distribution. In this case, we shall not look up the same values of n and p used in the last paragraph; we shall treat n' and p' simply as two parameters, using the prime to show that these are different numbers. If we replace n in (3) by sc and equate the means and variances of the match-

TABLE 2

Approximate Distributions

G-C is Gram-Charlier approximation; G-CB is the binomial modification of Gram-Charlier. Values of M(4, 13;h) and its 2d order G-C are taken from Greenwood (4), summed to four decimal places. The binomial approximation to M(4, 13;h) was chosen by method of text to have same mean and variance as exact distribution.

h	Exact M(8, 2;h)	Poisson $\nu = 2$	Binomial n=25;p=.08	4th Order G-C
0	1.00000	1.00000	1.00000	1.00000
1	.87335	.86466	.87564	.87330
2	.60334	.59399	.60528	.60340
3	.32369	.32332	.32317	.32361
4	.13676	.14288	.13509	.13671
5	.04640	.05265	.04514	.04643
6	.01285	.01656	.01229	.01287
7	.00295	.00435	.00277	.00295
8	.00056	.00110	.00052	.00055
9	.00009	.00024	.00008	.00009
10	.00001	.00005	.00001	.00001

h	Exact M(8, 3;h)	Poisson $\nu = 3$	Binomial n=30;p=.10	4th Order G-CB
0	1.00000	1.00000	1.00000	1.00000
1	.95645	.95021	.95761	.95646
2	.81402	.80085	.81630	.81400
3	.58715	.57681	.58865	.58712
4	.35295	.35277	.35256	.35295
5	.17702	.18474	.17549	.17704
6	.07467	.08392	.07319	.07468
7	.02674	.03351	.02583	.02674
8	.00820	.01191	.00778	.00820
9	.00217	.00380	.00202	.00217
10	.00050	.00110	.00045	.00050
11	.00010	.00029	.00009	.00010
12	.00002	.00007	.00002	.00002

h	Exact M(4, 13;h)	2nd Order G-C	Binomial n=68;p=1/17	Normal
0	1.00000	1.00000	1.00000	.990
1	.9838	.9839	.9839	.964
2	.9149	.9149	.9149	.902
3	.7707	.7705	.7708	.780
4	.5725	.5723	.5726	.602
5	.3712	.3712	.3713	.398
6	.2099	.2103	.2102	.210
7	.1047	.1045	.1045	.098

ing problem distribution and the binomial,

$$\nu = c = n'p'$$

$$\mu_2 = c^2(s - 1)/(sc - 1) = n'p'(1 - p'),$$

a pair of simultaneous equations which we can solve for n' and p' in terms of

s and c. For solutions

$$n' = c(sc - 1)/(c - 1) \tag{4}$$
$$p' = (c - 1)/(sc - 1).$$

Note that n' is not the size of the deck but is just a parameter for which we have solved. It may or may not be an integer. If it is, and p' is a value which can be found in a table, then the binomial distribution with parameters n' and p' will have the same mean and same variance as the matching problem distribution with s suits of c cards each. If n' is not an integer, the integer nearest to n' can be used, call it n'', and using $p'' = c/n''$ gives a binomial with the correct mean and very nearly the correct variance. The writer would guess, based on trying a few cases, that for decks larger than 25, one should be able to fit a binomial by this method to about three decimal places. Table 2 shows some examples of the use of the binomial as an approximation.

This approach suggests looking for known, or easily calculated, distributions which have the first two, three, or more moments in common with the matching problem distribution. Greenwood (4), following a suggestion by Mantel, used the Gram-Charlier series type B as an approximation. The Gram-Charlier uses the Poisson as a first approximation to the desired distribution; suitable multiples of the first, second, and successive differences of the Poisson are added to correct for the mean, second moment, etc. If $p(h)$ is the Poisson probability of h, then the first difference, $\Delta p(h)$, is defined to be $p(h) - p(h - 1)$. The second difference, $\Delta^2 p(h)$, is the first difference of the first difference, i.e., $\Delta p(h) - \Delta p(h - 1)$, and so on for higher differences. The reader will note that $\Delta p(0)$ is not well defined yet, since we must know what value to use for $p(-1)$—this is not ordinarily found in a table of the Poisson distribution. For purposes of the Gram-Charlier series, $p(h)$ is defined and equal to zero for all negative values of h. We are now ready to define the Gram-Charlier series. Where $m(h)$ is the distribution being approximated,

$$m(h) = p(h) + a_1\Delta p(h) + a_2\Delta^2 p(h) + a_3\Delta^3 p(h) + \cdots . \tag{5}$$

In principle, this is an infinite series; in practice, only the first few terms are used. The constants a_i appearing in (5) are determined by the moments of the matching problem and of the particular Poisson distribution used. The Poisson is usually chosen so that it has the same mean as the distribution being approximated; this makes the value of a_1 zero. If μ_i is the ith moment about the mean of the matching problem distribution and m_i the corresponding moment for the Poisson, the next three constants for the Gram-Charlier type B are given by

$$\begin{aligned} 2a_2 &= \mu_2 - m_2 \\ 6a_3 &= 6a_2 - (\mu_3 - m_3) \\ 24a_4 &= 36a_3 - 2a_2(6m_2 + 7) + (\mu_4 - m_4). \end{aligned} \tag{6}$$

As an example, suppose we wish to find an approximate value for $m(7, 2; 3)$, the probability of exactly three matches when both decks have seven suits of two cards per suit. (From Table 1 we see that the correct value is .32383 − .13580 = 0.18803.) As a first approximation, we look up the Poisson with mean 2, and find $p(3) = 0.18045$. This value is 0.00758 too low. For the next approximation, we need to know the variances and the value of $\Delta^2 p(3)$:

$$\mu_2 = c(n - c)/(n - 1) = 2(14 - 2)/(14 - 1) = 24/13$$

$$m_2 = 2$$

$$a_2 = \tfrac{1}{2}(24/13 - 2) = -1/13$$

$$\Delta p(3) = p(3) - p(2) = 0.18045 - 0.27067 = -0.09022$$

$$\Delta p(2) = p(2) - p(1) = 0.27067 - 0.27067 = 0.00000$$

$$\Delta^2 p(3) = \Delta p(3) - \Delta p(2) = -0.09022$$

$$a_2\Delta^2 p(3) = (-1/13)(-0.09022) = 0.00694.$$

Therefore, we have as a second approximation $p(3) + a_2\Delta^2 p(3) = 0.18739$, which is only 0.00064 too low. If we wished a still closer approximation, we could calculate the terms involving a_3 and a_4. An example of a fourth-order approximation is given in Table 2.

If the bionomial is used as a first approximation, higher-order approximations may be obtained by a modification of the Gram-Charlier series which uses the binomial instead of the Poisson. The formulas (5) and (6) are valid for this modification, provided that we re-interpret $p(h)$ and m_i to be the probability and moments of the binomial. It has been the writer's experience that the Poisson works well as a starting distribution when c is 2 or 3, and that the binomial gives a better fitting curve when c is more than 3. For decks of 24 or more cards, when all suits have the same number of cards, truncating the Gram-Charlier series (or the binomial modification) after terms which involve a_4 gives an approximation correct to about 0.00003 or less. For a more detailed description of the Gram-Charlier series and a derivation of the constants involved see Rietz (**10**).

More Complicated Decks

If the target deck does not have the same composition as the call deck, or if the suits do not all have the same number of cards, the estimation of significance levels may be more complicated. In many cases, the first complication can be avoided by the simple device of making sure the person who is attempting the classification knows what the categories are and how many objects are in each category. When it is possible to design the experiment so that this is done, several advantages are gained. The calculations for each

subject or trial are simpler for identical decks. Also, if each subject selects his own categories, there may be a different set of calculations necessary for each subject; whereas if all subjects use the same deck, only one distribution need be calculated. Finally, while the writer knows of no investigations of the power of this type of test, it seems clear that the subject has a better chance of showing his ability, if he has any, when both decks are the same. Since, in using the matching problem technique, the calculations are based on the decks actually used, there is no way to give credit for choosing the proper categories.

The calculations are also somewhat more complicated if different categories contain different numbers of objects. However, in many cases this is not a part of the experimental design which can be changed. For small decks, the exact distribution may have to be calculated. If so, the following procedure, taken from Greville (6), may be used. (This also works for calculating non-identical decks.) Suppose the two decks have the following composition:

	Suit	1	2	3	$\cdots$	s	Total
Cards Per Suit	Call Deck	m_1	m_2	m_3		m_s	n
	Target Deck	n_1	n_2	n_3		n_s	n

(7)

Let M_i be the smaller of m_i and n_i for each $i = 1, 2, \cdots, s$. Then the probability of exactly h matches, $m(h)$, is given by

$$m(h) = \frac{1}{n!} \sum_{j=h}^{n} (-1)^{j-h} \binom{j}{h} (n-j)! H_j , \tag{8}$$

where H_j is the coefficient of x^j in the expansion of

$$\prod_{i=1}^{s} \left[\sum_{k=0}^{M_i} \frac{m_i! n_i! x^k}{k!(m_i - k)!(n_i - k)!} \right].$$

As an example, suppose we wish to calculate the distribution of matches in the case of the handwriting expert given at the beginning of the paper. The two decks have the composition:

	Suit	1	2	3	Total
Cards Per Suit	Call Deck	5	2	3	10
	Target Deck	2	4	4	10

(9)

We notice one peculiar feature of non-identical decks at this point. Although both decks have 10 cards each, only 7 matches are possible: 2 in the first

Bureau Ednl. Research
.G COLLEGE
Dated

suit, 2 in the second, and 3 in the third. Thus, we will expect to get non-zero values of $m(h)$ only for $h = 0, 1, \cdots, 7$.

First expand the generating function (9) to get H_i :

$$(1 + 10x + 20x^2)(1 + 8x + 12x^2)(1 + 12x + 36x^2 + 24x^3)$$
$$= 1 + 30x + 364x^2 + 2296x^3 + 8064x^4 + 15648x^5 + 15360x^6 + 5760x^7.$$

Then calculate the factor $Q_j = (n - j)!H_j$ for each $j = 0, 1, \cdots, 7$. Next calculate $n!m(h) = \sum_{j=h}^{7} (-1)^{j-h}\binom{j}{h}Q_j$, for each $h = 0, 1, \cdots, 7$, where $\binom{j}{h}$ is a binomial coefficient, which can be obtained from tables. With a table of binomial coefficients and the values of Q_j , which we have calculated, this sum can be obtained rather quickly on a desk calculator. Finally, divide each value by $n!$ to get $m(h)$, the probability of exactly h matches. In order to get the probability of h or more matches, add $\sum_{i=h}^{7} m(i)$, getting the values listed in Table 3.

TABLE 3

Handwriting Expert Example

h	Exact	Normal	Binomial
0	1.00000	0.9922	1.00000
1	.96984	.9582	.9718
2	.84762	.8508	.8507
3	.62064	.6368	.6172
4	.35556	.3632	.3504
5	.15238	.1492	.1503
6	.04444	.0418	.0474
7	.00952	.0078	.0106

To approximate this distribution, using either the normal or the binomial, first calculate the mean and variance. For this purpose use a formula derived by Battin (2). (Battin gives a brief review of the mathematical literature on the matching problem prior to 1942 and lists an extensive bibilography.) He uses the technique of generating functions to get formulas applicable to decks of arbitrary composition. He also extends the matching problem to more than two decks, with the possibilities of two-card matches, three-card matches, etc. Here, only his results for two decks are used.

If the two decks have the composition specified in (7) above, then the mean number of matches, ν, and the variance, σ^2, are given by:

$$\nu = \frac{1}{n} \sum m_i n_i \tag{10}$$

$$\sigma^2 = \frac{1}{n^2(n-1)} [(\sum m_i n_i)^2 - n \sum m_i n_i (m_i + n_i) + n^2 \sum m_i n_i].$$

These formulas have been rearranged slightly from the form in which Battin listed them, for convenience in calculating. The products $m_i n_i$ and $m_i n_i (m_i + n_i)$ can be obtained quickly from the table (7) giving the composition of the deck, and the rest of the calculation is straightforward. We shall outline the calculations for the handwriting example.

i	1	2	3	Total
m_i	5	2	3	10
n_i	2	4	4	10
$m_i n_i$	10	8	12	30
$m_i + n_i$	7	6	7	
$m_i n_i (m_i + n_i)$	70	48	84	202

$$\nu = 30/10 = 3$$

$$\sigma^2 = [1/(100)(9)][(30)^2 - 10(202) + 100(30)]$$

$$\sigma^2 = 1880/900 = 2.0889$$

$$\sigma = \sqrt{2.0889} = 1.45.$$

Since the normal is a continuous distribution, while the matching problem distribution is discrete with probability "concentrated" on the integers, we must make a correction when using the normal as an approximation. That is, instead of the probability that $X = h$, we want the probability that X is between $h - 1/2$ and $h + 1/2$; if we want the probability of h or more matches, we must find $\text{Prob}(X \geq h - 1/2)$, where X is normally distributed with mean 3 and standard deviation 1.45. (In the special case where each deck has only two suits, the probability is concentrated on the even integers, and we must subtract 1 instead of 1/2 in order to get the proper correction.) Since we have available a table of the distribution of Y, a normal variate with zero mean and unit standard deviation, we look up

$$\text{Prob}\left(Y \geq \frac{h - \frac{1}{2} - 3}{1.45}\right).$$

The values found for $h = 0, \cdots, 7$ are listed in Table 3.

The last entry in Table 3 gives the binomial approximation to our matching problem. The binomial which came closest to fitting was that for which $n = 10$ and $p = .3$, which has mean 3 and variance 2.10.

REFERENCES

1. Anderson, T. W. On card matching. *Ann. math. Statist.*, 1943, 14, 426-435.
2. Battin, I. L. On the problem of multiple matching. *Ann. math. Statist.*, 1942, 13, 294-305.
3. Feller, W. An introduction to probability theory and its applications, Volume 1. New York: Wiley, 1950. pp. 62-69.
4. Greenwood, R. E. Probabilities of certain solitaire card games. *J. Amer. statist. Ass.*, 1953, 48, 88-93.
5. Greville, T. N. E. Exact probabilities for the matching hypothesis. *J. Parapsychology*, 1938, 2, 55-59.
6. Greville, T. N. E. The frequency distribution of a general matching problem. *Ann. math. Statist.*, 1941, 12, 350-354.
7. Hamilton, C. H. Comment on question 8: variance of number of matches. *Amer. Statistician*, June, 1948, 2, 23.
8. Huntington, E. V. Exact probabilities in certain card matching problems. *Science*, 1937, 86, 499-500.
9. Olds, E. G. A moment-generating function which is useful in solving certain matching problems. *Bull. Amer. math. Soc.*, 1938, 44, 407-413.
10. Rietz, H. L. Mathematical statistics, pp. 60-76 and 156-172. Chicago: Open Court Publishing Co., 1927.
11. U. S. Department of Commerce, National Bureau of Standards. Tables of the binomial probability distribution. Washington: U.S. Government Printing Office, 1949.

Manuscript received 1/10/55

Revised manuscript received 8/15/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

A COMPARISON OF GAME THEORY AND LEARNING THEORY

HERBERT A. SIMON
CARNEGIE INSTITUTE OF TECHNOLOGY

It is shown that Estes' formula for the asymptotic behavior of a subject under conditions of partial reinforcement can be derived from the assumption that the subject is behaving rationally in a certain game-theoretic sense and attempting to minimax his regret. This result illustrates the need for specifying the frame of reference or set of the subject when using the assumption of rationality to predict his behavior.

Learning theory and game theory (together with the closely related statistical decision theory) purport to provide theories of *rational* behavior. Implicit in any theory of learning is a motivational assumption that learning consists in the acquisition of a pattern of behavior appropriate to *goal achievement*, *need reduction*, or the like. In parallel fashion game theory and statistical decision theory are concerned with discovering the course of action in a particular situation that will optimize the attainment of some objective pay-off.

In order to gain a better understanding of the concepts of *rationality* underlying these two bodies of theory, it would be interesting to construct a situation in which predictions made from these theories could be compared and then checked against experimental data on actual behavior. One situation of this kind received considerable attention at the Santa Monica Conference on Decision Processes (**2**, **3**, **4**). The experiment is one involving partial reinforcement. At each trial the subject chooses between two alternatives. Each alternative is rewarded on a certain per cent of the trials in which it is chosen (the trials rewarded being randomly determined) and not rewarded on the remaining trials in which it is chosen; the per cent of rewarded trials is in general different for the two alternatives. The learning theory advanced by Estes provides a prediction as to the frequency (in the limit as the number of trials increases) with which the first alternative will be chosen in preference to the second (**2**). The same frequency is predicted by the Bush-Mosteller theory when certain assumptions of symmetry are made with respect to the parameters that appear in their model (**1**, ch. 8). Estes reports several experiments that confirm predictions from his theory.

When this experimental situation was described to a number of game theorists at the Santa Monica conference, they pointed out that a *rational* individual would first estimate, by experimenting, which of the two alternatives had the greatest probability of reward, and would subsequently always

select that alternative which would not be predicted by the Estes theory. Flood has defended the choices predicted by the Estes theory against the charge of irrationality, basing his defense on two points (4, p. 288):

> (a) The proper definition of payoff utilities would be unclear in attempts to apply game-theoretic arguments to a real case, and there is a reasonable payoff matrix that would rationalize the reported behavior. Thus, if the organism's object were to maximize its score rather than its expectation, then it should sometimes not tend to use a pure strategy . . .
>
> (b) The von Neumann-Morgenstern game theory is inapplicable in this situation unless the organism can assume safely that the experimental stimulus is generated by a stationary stochastic process. For example, if the organism believes that there may be some pattern (non-stationarity) over time, in the stimulus, then it can often do better by using a mixed strategy rather than a pure one, for the latter would give it no way to discover any pattern effect.

In the next section, by combining in an appropriate fashion the two considerations advanced by Flood—that is, by assuming (a) the subject is trying to maximize something other than expected payoff, and (b) the subject does not believe or expect that the probability of reward from each alternative is fixed—it will be shown that the behavior predicted by the Estes theory and actually observed in experiments is rational in the sense of game theory (or at least in one of the many senses consistent with game theory). In a final section, the implications of this result will be discussed briefly.

Game-Theoretical Derivation of Estes' Result

Consider a partial reinforcement experiment in which there are two alternatives of behavior, A_1 and A_2. If A_1 is chosen on a particular trial, it is rewarded with probability π_1; if A_2 is chosen, it is rewarded with probability π_2. Let $p_1(t)$ be the probability that the subject chooses A_1 on the tth trial; $p_2(t) = (1 - p_1)$ the probability that he chooses A_2. From the postulates of his learning model, Estes (2) predicts that the asymptotic value of p_1 as the number of trials increases will be p_1^*,

$$p_1^* = \frac{1 - \pi_2}{(1 - \pi_1) + (1 - \pi_2)}. \tag{1}$$

This value for p_1^* may be obtained as the steady state of the stochastic process

$$\tilde{p}(t + 1) = \Pi\tilde{p}(t), \quad \text{where} \quad \Pi = \begin{pmatrix} \pi_1 & (1 - \pi_2) \\ (1 - \pi_1) & \pi_2 \end{pmatrix}; \tag{2}$$

for

$$p_1(t + 1) = \pi_1 p_1(t) + (1 - \pi_2)[1 - p_1(t)], \tag{3}$$

so that, if $p_1(t+1) = p_1(t) = p_1^*$,

$$(1 - \pi_1 + 1 - \pi_2)p_1^* = (1 - \pi_2), \tag{4}$$

from which (1) follows immediately.

We see that in Estes' theory π_1 is the probability that A_1 will be rewarded; but it is also the asymptotic probability that, having chosen A_1 on a given trial, the subject will choose it again on the next succeeding trial. A similar interpretation can be given to π_2. Hence, we may interpret π_1 and π_2 as the conditional probabilities of *persistent* behavior when the subject has just chosen A_1 or A_2, respectively; while $(1 - \pi_1)$ and $(1 - \pi_2)$ are the corresponding conditional probabilities of a *shift* in behavior.

For specificity, let us consider the case where $\pi_1 > \pi_2$. Then the game-theoretical objection to regarding as rational the asymptotic behavior predicted by the Estes model is that the subject could increase his expected reward by always choosing A_1. For then the expected reward would be

$$\pi_1 > p_1^*\pi_1 + (1 - p_1^*)\pi_2, \tag{5}$$

where the terms on the right-hand side of the inequality are easily seen to be the expected reward for the Estes model.

But the rationality of this game-theoretical solution is compelling only under the assumption that the reward probabilities are known to the subject, and known to be constant. These are the assumptions that Flood challenges. Let us consider an alternative set of assumptions which, while not the only possible such set, has some plausibility.

(*i*) The subject takes as given and fixed the π corresponding to the alternative he has chosen on the last trial. That is, he assumes the probability of reward to be π_1 or π_2, if he persists in choosing again A_1 or A_2, as the case may be.

(*ii*) The subject expects that if he *shifts* from the alternative just chosen to the other one, the probability of reward is unknown and dependent on a strategy of nature.

(*iii*) The subject does not wish to persist in his present behavior if there is a good chance of increased reward from shifting. He measures his success on each trial not from the reward received, but from the difference between the reward actually received and the reward that *could* have been attained if he had outguessed nature. In the terminology of L. J. Savage, he wishes to minimize his *regret*.

We may formalize these assumptions as follows: On each trial, the subject chooses between (*i*) *persisting* or (*ii*) *shifting* his choice. If he persists, he is rewarded with probability π (where $\pi = \pi_1$, or $\pi = \pi_2$ depending on whether the previous choice was A_1 or A_2, respectively), irrespective of the strategy adopted by nature. If he shifts, he will receive a reward of 0 if nature adopts her strategy (α), and a reward of 1 if nature adopts her strategy

select that alternative which would not be predicted by the Estes theory. Flood has defended the choices predicted by the Estes theory against the charge of irrationality, basing his defense on two points (4, p. 288):

> (a) The proper definition of payoff utilities would be unclear in attempts to apply game-theoretic arguments to a real case, and there is a reasonable payoff matrix that would rationalize the reported behavior. Thus, if the organism's object were to maximize its score rather than its expectation, then it should sometimes not tend to use a pure strategy . . .
>
> (b) The von Neumann-Morgenstern game theory is inapplicable in this situation unless the organism can assume safely that the experimental stimulus is generated by a stationary stochastic process. For example, if the organism believes that there may be some pattern (non-stationarity) over time, in the stimulus, then it can often do better by using a mixed strategy rather than a pure one, for the latter would give it no way to discover any pattern effect.

In the next section, by combining in an appropriate fashion the two considerations advanced by Flood—that is, by assuming (a) the subject is trying to maximize something other than expected payoff, and (b) the subject does not believe or expect that the probability of reward from each alternative is fixed—it will be shown that the behavior predicted by the Estes theory and actually observed in experiments is rational in the sense of game theory (or at least in one of the many senses consistent with game theory). In a final section, the implications of this result will be discussed briefly.

Game-Theoretical Derivation of Estes' Result

Consider a partial reinforcement experiment in which there are two alternatives of behavior, A_1 and A_2. If A_1 is chosen on a particular trial, it is rewarded with probability π_1; if A_2 is chosen, it is rewarded with probability π_2. Let $p_1(t)$ be the probability that the subject chooses A_1 on the tth trial; $p_2(t) = (1 - p_1)$ the probability that he chooses A_2. From the postulates of his learning model, Estes (2) predicts that the asymptotic value of p_1 as the number of trials increases will be p_1^*,

$$p_1^* = \frac{1 - \pi_2}{(1 - \pi_1) + (1 - \pi_2)}. \tag{1}$$

This value for p_1^* may be obtained as the steady state of the stochastic process

$$\bar{p}(t + 1) = \Pi\bar{p}(t), \quad \text{where} \quad \Pi = \begin{pmatrix} \pi_1 & (1 - \pi_2) \\ (1 - \pi_1) & \pi_2 \end{pmatrix}; \tag{2}$$

for

$$p_1(t + 1) = \pi_1 p_1(t) + (1 - \pi_2)[1 - p_1(t)], \tag{3}$$

so that, if $p_1(t+1) = p_1(t) = p_1^*$,

$$(1 - \pi_1 + 1 - \pi_2)p_1^* = (1 - \pi_2), \tag{4}$$

from which (1) follows immediately.

We see that in Estes' theory π_1 is the probability that A_1 will be rewarded; but it is also the asymptotic probability that, having chosen A_1 on a given trial, the subject will choose it again on the next succeeding trial. A similar interpretation can be given to π_2. Hence, we may interpret π_1 and π_2 as the conditional probabilities of *persistent* behavior when the subject has just chosen A_1 or A_2, respectively; while $(1 - \pi_1)$ and $(1 - \pi_2)$ are the corresponding conditional probabilities of a *shift* in behavior.

For specificity, let us consider the case where $\pi_1 > \pi_2$. Then the game-theoretical objection to regarding as rational the asymptotic behavior predicted by the Estes model is that the subject could increase his expected reward by always choosing A_1. For then the expected reward would be

$$\pi_1 > p_1^*\pi_1 + (1 - p_1^*)\pi_2, \tag{5}$$

where the terms on the right-hand side of the inequality are easily seen to be the expected reward for the Estes model.

But the rationality of this game-theoretical solution is compelling only under the assumption that the reward probabilities are known to the subject, and known to be constant. These are the assumptions that Flood challenges. Let us consider an alternative set of assumptions which, while not the only possible such set, has some plausibility.

(*i*) The subject takes as given and fixed the π corresponding to the alternative he has chosen on the last trial. That is, he assumes the probability of reward to be π_1 or π_2, if he persists in choosing again A_1 or A_2, as the case may be.

(*ii*) The subject expects that if he *shifts* from the alternative just chosen to the other one, the probability of reward is unknown and dependent on a strategy of nature.

(*iii*) The subject does not wish to persist in his present behavior if there is a good chance of increased reward from shifting. He measures his success on each trial not from the reward received, but from the difference between the reward actually received and the reward that *could* have been attained if he had outguessed nature. In the terminology of L. J. Savage, he wishes to minimize his *regret*.

We may formalize these assumptions as follows: On each trial, the subject chooses between (*i*) *persisting* or (*ii*) *shifting* his choice. If he persists, he is rewarded with probability π (where $\pi = \pi_1$, or $\pi = \pi_2$ depending on whether the previous choice was A_1 or A_2, respectively), irrespective of the strategy adopted by nature. If he shifts, he will receive a reward of 0 if nature adopts her strategy (α), and a reward of 1 if nature adopts her strategy

(β). The payoff matrix corresponding to these assumptions is:

	(α)	(β)
(i)	π	π
(ii)	0	1

where rows correspond to the subject's strategies and columns to nature's strategies. *Regret* is defined as the difference between the actual payoff for a given pair of strategies [e.g. (i), (β)], and the payoff that *could* have been realized, if the strategy actually employed by nature had been anticipated [e.g., (ii), (β)]. Performing the indicated subtractions, the regret matrix is:

	(α)	(β)
(i)	0	$(\pi - 1)$
(ii)	$-\pi$	0

(This was obtained from the first matrix by subtracting from each element the largest element in the same column).

Now let ρ be the probability that the subject uses strategy (i), i.e., persists, μ be the probability that nature uses strategy (α). Then the expected regret will be

$$\begin{aligned} R &= \rho\mu\cdot 0 + \rho(1-\mu)(\pi - 1) + (1-\rho)\mu(-\pi) + (1-\rho)(1-\mu)\cdot 0 \\ &= \rho(1-\mu)(\pi - 1) - (1-\rho)\mu\pi. \end{aligned} \tag{6}$$

The conditions that the regret be minimum (strictly, *minimax*) are given by

$$\frac{\partial R}{\partial \rho} = \frac{\partial R}{\partial \mu} = 0. \tag{7}$$

Using the second of these equations, we obtain from (6)

$$0 = -\rho(\pi - 1) - (1-\rho)\pi, \tag{8}$$

whence

$$\rho = \pi. \tag{9}$$

Hence the subject would persist with probability π and shift with probability $(1 - \pi)$. But this is precisely the postulate contained in (2).

Hence, we have shown that the behavior predicted by Estes' theory is identical with that which would be exhibited by a *rational* subject intent on *minimaxing* his regret.

Comments on the Derivation

We need not try to decide whether the subjects who behave in conformity with the predictions of Estes' theory are minimaxing regret, or whether they are simply behaving in the adaptive fashion implied by the usual learning mechanisms. Most economists and statisticians would be tempted to accept the former interpretation, most psychologists the latter. It is not immediately obvious what source, other than introspection, would provide evidence for deciding the issue.

Perhaps the most useful lesson to be learned from the derivation is the necessity for careful distinctions between *subjective* rationality (i.e., behavior that is rational, given the perceptual and evaluational premises of the subject), and *objective* rationality (behavior that is rational as viewed by the experimenter). Because this distinction has seldom been made explicitly by economists and statisticians in their formulations of the problem of rational choice, considerable caution must be exercised in employing those formulations in the explanation of observed behavior.

To the experimenter who knows that the rewards attached to the two behaviors A_1 and A_2 are random, with constant probabilities, it appears unreasonable that the subject should not learn to behave in such a way as to maximize this expected gain—always to choose A_1. To the subject, who perceives the situation as one in which the probabilities may change, and who is more intent on outwitting the experimenter (or nature) than on maximizing expected gain, rationality is something quite different. If rationality is to have any meaning independent of the perceptions of the subject we must distinguish between the rationality of the perceptions themselves (i.e., whether or not the situation as perceived is the real situation) and the rationality of the choice, given the perceptions.

If we accept the proposition that organismic behavior may be subjectively rational but is unlikely to be objectively rational in a complex world then the postulate of rationality loses much of its power for predicting behavior. To predict how economic man will behave we need to know not only that he is rational, but also how he perceives the world—what alternatives he sees, and what consequences he attaches to them (**5**). We should not jump to the conclusion, however, that we can therefore get along without the concept of rationality. While the Estes model predicts the behavior of naive subjects under partial reinforcement, we observe (**3**) that persons trained in game theory and placed in the same situation generally learn to choose A_1 consistently. It appears simpler to postulate here a change in set—a change in the perceptual model—rather than to attempt to explain this

behavior in terms of simpler learning-theoretic models. If anything was learned during the series of trials by the subjects who were game theorists, it was the appropriate perceptual model and not the appropriate behavior once that model is assumed.

REFERENCES

1. Thrall, R. M., Coombs, C. H., and Davis, R. L. (Eds.) Decision processes. New York: Wiley, 1954.
2. Estes, W. K. Individual behavior in uncertain situations: an interpretation in terms of statistical association theory. In (1), chap. 9.
3. Flood, M. W. Game-learning theory and some decision-making experiments. In (1), chap. 10.
4. Flood, M. W. Environmental non-stationarity in a sequential decision-making experiment. In (1), chap. 18.
5. Simon, H. A. A behavorial model of rational choice. *Quart. J. Econ.*, 69, 1955, 99–118.

Manuscript received 4/18/55

Revised manuscript received 5/28/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

"BEST POSSIBLE" SYSTEMATIC ESTIMATES OF COMMUNALITIES*

LOUIS GUTTMAN

THE ISRAEL INSTITUTE OF APPLIED SOCIAL RESEARCH

At least four approaches have been used to estimate communalities that will leave an observed correlation matrix R Gramian and with minimum rank. It has long been known that the square of the observed multiple-correlation coefficient is a lower bound to any communality of a variable of R. This lower bound actually provides a "best possible" estimate in several senses. Furthermore, under certain conditions basic to the Spearman-Thurstone common-factor theory, the bound must equal the communality in the limit as the number of observed variables increases. Otherwise, this type of theory cannot hold for R.

I. *Introduction*

One of the intriguing problems of factor analysis has been to find a formula for communalities that will minimize the rank of an arbitrary correlation matrix R. More explicitly, the problem is to find a diagonal matrix U such that $R - U^2$ is Gramian and of minimum rank.

Let n denote the order of R (and of U), and m the minimum rank for Gramian $R - U^2$. At least four approaches have been used to estimate communalities that will yield m:

(a) trial-and-error exact formulas
(b) exact formulas for special cases of R
(c) successive approximations
(d) lower bounds.

The main thesis of this paper is that, in certain senses, the last-mentioned of these four approaches provides "best possible" estimates of communalities for an arbitrary R, even though biased in general by being underestimates.

Let u_j be the jth diagonal element of any U that leaves $R - U^2$ Gramian (whether or not with minimum rank), and let h_j^2 be the corresponding communality:

$$h_j^2 = 1 - u_j^2 \qquad (j = 1, 2, \cdots, n). \tag{1}$$

Let ρ_j denote the multiple correlation coefficient of the jth variable in R on the remaining $n - 1$ variables, and σ_j the corresponding standard error

*This research was facilitated by a grant from the Lucius N. Littauer Foundation to the American Committee for Social Research in Israel in order to promote methodological work of the Israel Institute of Applied Social Research.

of estimate (assuming all observed variables to have unit variances):

$$\rho_j^2 = 1 - \sigma_j^2 \qquad (j = 1, 2, \cdots, n). \tag{2}$$

Then it has been shown (2, 92*f*; 3, 293) that always

$$\rho_j^2 \leqq h_j^2 \qquad (j = 1, 2, \cdots, n). \tag{3}$$

No better general lower bound to h_j^2 has yet been established than ρ_j^2.

We shall prove here that there exist many nonsingular matrices R for which the equality in (3) holds for $n - m$ of the minimizing communalities—all but m of the ρ_j^2 are actual communalities (the remaining communalities equal unity). Such matrices R, however, are of a very restricted type.

A more generally useful result that we shall establish applies to the typical R postulated in the Spearman-Thurstone theory. This school of thought believes a common-factor analysis is meaningful only if m is small compared with n. We shall prove that if the ratio of m to n tends to zero as $n \to \infty$, then all except possibly a zero proportion of the ρ_j^2 must tend to the rank-minimizing h_j^2. *If the Spearman-Thurstone hypothesis is correct for a given R, then the ρ_j^2 must almost always be very good approximations to the h_j^2 when n is large.* (Conversely, if the approximation is bad for many ρ_j, then the Spearman-Thurstone hypothesis of a limited number of common factors must be false.)

An even more general result refers to all R, regardless of the ratio of m to n. If there is to be one and only one unique-factor variable that can yield the uniqueness u_j^2, then it must be that the limit of σ_j^2 must be u_j^2 as $n \to \infty$ (or it must be that $\rho_j^2 \to h_j^2$). Conversely, if σ_j^2 does not tend to u_j^2 as $n \to \infty$, then there is more than one "unique" variable that can provide the same loading u_j^2 (and satisfy all other algebraic requirements of common-factor theory); the larger the difference between σ_j^2 and u_j^2 (or between ρ_j^2 and h_j^2), the larger the possible difference between alternative "unique" parts for the same jth observed variable of R.

Other important properties of the lower bounds ρ_j^2 will be established. Before going on to our new results, it may be helpful to review briefly the four approaches listed above.

(*a*) *Trial and Error*

Assuming that sampling error and rounding-off errors in computations are nonexistent, trial and error is bound to yield an exact numerical answer when $m < n/2$; the diagonal elements of U^2 in such cases are *rational functions* of the non-diagonal elements of R (cf. 8). It may turn out, of course, that U^2 is not uniquely determined; two or more different U^2 for the same R may yield m in many cases. When $m \geqq n/2$, trial and error can lead again to an expression for each communality, although in non-rational form in general. Again, multiple solutions for minimizing U^2 may occur.

(b) *Special Exact Formulas*

Some special cases of R make possible exact and rational formulas that need no apparent resort to trial and error. The known cases are for $m < n/2$, the most celebrated being Spearman's where $m = 1$. Thurstone has summarized a number of such formulas (**8**, ch. XIII). A caution should be added to Thurstone's discussion to the effect that not all the apparent solutions may yield Gramian U^2 nor leave $R - U^2$ Gramian. Actually these formulas beg the question, for it is generally not known in advance whether or not $m < n/2$. A specialized formula in effect must be tried on the given R to see if it works. Use of specialized formulas thus seems to be but a modified type of trial and error.

(c) *Successive Approximations*

Attempts have been made to avoid a direct exact solution for U^2 by taking recourse instead to successive approximations. An approximation U_1^2 is guessed, and $R - U_1^2$ is "factored" until residuals are considered small enough, leading to a second approximation U_2^2, etc. It has been claimed that such a procedure generally converges to a satisfactory U^2 (cf. **8**, p. 295). Algebraic proof of such convergence has never been published to our knowledge. For many iterative processes, the value to which convergence takes place depends on the initial trial value. That this may be the case for the above procedure seems evident when one recalls that there are many correlation matrices which do not have a unique set of communalities. Also, unless proof is given to the contrary, there is no reason to believe that successive approximations may not converge to some U where $R - U^2$ is *not* of minimum rank, if convergence takes place at all.

The issue of successive approximations is further beclouded by sampling considerations. Lawley's maximum likelihood solution seems the most appropriate put forward to date, as Rao points out (**7**). To attain precision in the sampling theory, apparently some restrictions have been introduced as to the nature of the population R, else the possibility of equally minimizing alternative solutions would remain. Again, it is not clear when a given R obeys these restrictions or when the sampling theory is valid in practice. [After the above was written, the writer received a copy of reference (**1**) in which a numerical example is given of the failure of Lawley's iterative procedure to converge properly.]

(d) *Lower Bounds*

If we again ignore sampling and rounding-off errors, it is always possible to establish useful lower bounds to communalities without any trial and error and without any hypothesis about or restrictions on R. The best of the lower bounds thus far established are the ρ_i^2, according to inequality (3)

above. It is often more convenient to discuss uniqueness rather than communalities, or to use inequality (4) rather than (3):

$$\sigma_j^2 \geqq u_j^2 \qquad (j = 1, 2, \cdots, n). \tag{4}$$

An important feature of the bounds in (3) and (4) is that they hold whether or not there is a multiple solution for U^2; they hold for all possible solutions simultaneously. Indeed, they lead to a criterion for choosing among multiple solutions, as indicated in the next section.

II. *Relationship to the Determinacy of Unique-Factor Scores*

Let r_j denote the multiple-correlation coefficient on the n observed variables of a unique-factor variable hypothesized to yield the uniqueness u_j^2. It has been shown in (6) that

$$r_j^2 = \frac{u_j^2}{\sigma_j^2} \qquad (j = 1, 2, \cdots, n). \tag{5}$$

Since (5) holds for all solutions U^2, it suggests that when a choice is necessary that which makes the inequalities (4) as small as possible is most desirable; the denominator on the right of (5) is fixed for j, so that such a choice makes the individual scores on the unique factor as determinate as possible from the observed data, or the r_j^2 as close as possible to unity. It has been shown that this also often tends to make individual scores on the *common-factor* variables as determinate as possible (6).

Should the approximations (4) for U^2 turn out *not* to be close in a given case, then the factor analysis itself may be regarded as not very useful or definitive. For it has been shown in (6) that determining factor *loadings* alone—common and unique—can be far from sufficient for pinning down *scores* on the hypothesized factors. *Alternative sets of scores for a given hypothesized factor can exist which yield identical loadings and yet correlate negligibly with each other*, according to formula (6),

$$r_j^* = 2r_j^2 - 1 \qquad (j = 1, 2, \cdots, n), \tag{6}$$

where r_j is given by (5) and r_j^* is the minimal correlation always attainable between two alternative sets of scores for the same unique factor hypothesized to underlie u_j^2. [According to (6), if $r_j^2 = .5$, then $r_j^* = 0$, or alternative score solutions for the same jth unique factor always exist that correlate *zero* with each other. Even if r_j^2 is as large as .9, this raises r_j^* only to .8. An equation parallel to (6) holds for common factors.]

III. *The "Best Possible" Estimates*

Can inequality (4) be improved on without recourse to some form of trial and error or use of specialized hypotheses? This does not seem possible. According to (5) this would imply some advance information on the r_j^2; there is no apparent way of getting such information on the r_j^2 in a

universally systematic manner. The situation seems to be the reverse: r_i^2 is determined by u_i^2 rather than conversely.

The rest of this paper will be devoted largely to showing that (4) is actually a "best possible" inequality in the sense that the phrase "best possible" is usually used mathematically for inequalities. The essential characteristics are that (a) many correlation matrices R exist for which the *equality* in (4) is actually attained at the same time that minimum rank m is attained, and (b) the inequalities in (4) must tend to equalities as n increases, under certain general conditions important to the theory of common-factor analysis. The bounds improve systematically in general as n increases, or as there is more information available from more observed variables. Furthermore, inequality (4) leads to inequalities for m that are also "best possible," and is closely related to the problem of estimating individual scores on the unique factors without any rank assumptions, via image analysis.

In virtually all attempts to solve the communality problem—whether exactly or by successive approximations—the problem is stated as for a fixed and finite n, or where R is from a finite number of n observed variables. It seems appropriate to ask also what happens to communalities as n increases or decreases.

While this issue is not discussed very explicitly by most writers, it usually seems implied that if the additional variables retain the same general kind of content as the initial ones, communalities of the initial ones should remain constant for all n sufficiently large. This would imply that for n small enough we should generally have $m > n/2$, or easy exact computations for U^2 (even ignoring sampling error) should be the exception rather than the rule. Having $m > n/2$ for relatively small n does not preclude m from remaining constant—and hence becoming relatively small—as n increases. It does imply that multiple solutions should be quite prevalent for finite n in practice. Furthermore, it cautions that an apparently exact solution for finite n may be but an artifact due to the finiteness of the number of variables observed.

It would be desirable, in view of all the preceding considerations, to have a systematic way of getting information about communalities with no assumptions whatsoever about R, yet without resorting to trial and error. Furthermore, this information should remain valid as n increases.

One of the virtues of the bounds (3) and (4) is that they possess these qualities in a simple and direct manner. This seems to be another type of "best possible" property from that usually considered, and one which appears peculiarly relevant to the problem of factor analysis.

IV. *Attaining Equality When n Is Finite*

If $\sigma_i^2 = 0$ for some j (so that $\rho_i^2 = 1$), then it must be that $u_i^2 = 0$ from (4) and the fact that a uniqueness cannot be negative. Here is one kind of

special circumstance wherein our bound becomes an exact estimate even when n is finite. In practice, this is not to be expected, since having one observed variable perfectly predictable from all the rest makes R singular.

Many cases of nonsingular R also exist for which the equality in (4) holds and n is finite. We shall exhibit some now. To this end, let us first recall that the σ_j^2 are the reciprocals of the corresponding main diagonals of R^{-1}. The following notation will be useful here and also later. Let S^{-2} (the inverse of S^2) be the diagonal matrix with the same main diagonal elements as R^{-1}. Then the jth main diagonal element of S^2 itself is simply σ_j^2 $(j = 1, 2, \cdots, n)$:

$$S^2 = [\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2]. \tag{7}$$

If R is nonsingular, there exists a nonsingular matrix F such that

$$R = FF'. \tag{8}$$

F can be chosen in infinitely many ways for (8) when $n \geqq 2$, but always we can rearrange variables to find an F of the form

$$F = \begin{Vmatrix} A & 0 \\ B & C \end{Vmatrix}, \tag{9}$$

where A is a nonsingular square submatrix of order m, B is of order $(n - m) \times m$, and C is nonsingular and of order $n - m$. From (8) and (9),

$$R = \begin{Vmatrix} AA' & AB' \\ BA' & BB' + CC' \end{Vmatrix}. \tag{10}$$

It is easily verified that the inverse of F is given by

$$F^{-1} = \begin{Vmatrix} A^{-1} & 0 \\ -C^{-1}BA^{-1} & C^{-1} \end{Vmatrix}. \tag{11}$$

From (8), $R^{-1} = (F^{-1})'F^{-1}$, or using (11)

$$R^{-1} = \begin{Vmatrix} G & H' \\ H & (CC')^{-1} \end{Vmatrix}, \tag{12}$$

where

$$G = (AA')^{-1} + (A^{-1})'B'(CC')^{-1}BA^{-1} \tag{13}$$

and

$$H = -(CC')^{-1}BA^{-1}. \tag{14}$$

Now consider the special case where CC' is a diagonal matrix. Then $(CC')^{-1}$ is also diagonal. According to (12) and (7), $(CC')^{-1}$ constitutes the

lower right-hand submatrix of S^{-2}, or CC' constitutes the corresponding submatrix of S^2 and defines the σ_j^2 for $j = m + 1, m + 2, \cdots, n$. If we subtract this submatrix CC' from the lower right-hand corner of R in (10), we are clearly left with a reduced R that is Gramian and of rank m, it being the product of $[A\ B]'$ and its transpose. Thus we have:

Theorem 1. *If R can be factored into an F of the form* (9) *where CC' is diagonal* (*and A and C are nonsingular*), *then the main diagonal elements of CC' are the respective σ_j^2 for $j = m + 1, m + 2, \cdots, n$. If these $n - m$ σ_j^2 in CC' are subtracted from the corresponding main diagonal elements of R, the resulting matrix will be Gramian and of rank m.*

According to Theorem 1, when m is the actual minimal rank possible for Gramian $R - U^2$, then the first m diagonal elements of U^2 can be set equal to zero, and the last $n - m$ diagonal elements equal to the corresponding σ_j^2 as given by CC'. Thus, the last $n - m$ of the σ_j^2 serve exactly as rank-minimizing uniquenesses, or the equality in (4) holds for $j = m + 1, m + 2, \cdots, n$.

Notice that the *first* m uniquenesses implied by Theorem 1 are zero and not equal to the σ_j^2. If the first m σ_j^2 were also subtracted out from the main diagonal of R, then the resulting $R - S^2$ would in general not be Gramian, nor of minimum rank (cf. **4**).

Theorem 1 holds even when the m in it is not minimal. It is always possible to use the Theorem for the case where C is of order 1, and hence CC' is *necessarily* a diagonal matrix. This provides:

Corollary. For any nonsingular R, if any one σ_j^2 is subtracted from the corresponding main diagonal element of R, then the resulting matrix is of rank $n - 1$.

This result was partly indicated by Thurstone in his discussion of the "diagonal" method of matrix factoring (better known to mathematicians as the Schmidt or Gram-Schmidt process of orthogonalization), but without noticing apparently that his implied uniqueness was exactly σ_n^2 (**8**, p. 308).

We have thus completed showing that there are many matrices for which many of the σ_j^2 can serve as rank-minimizing uniquenesses. Also, we have the curious result that any one of the σ_j^2 alone will reduce nonsingular R to a Gramian matrix of rank $n - 1$.

V. *Equality in the Limit as $n \to \infty$*

We have already seen in Part III how, if a "unique"-factor variable is really to be uniquely determined for a given u_j^2, then we must have $u_j^2/\sigma_j^2 \to 1$ as $n \to \infty$, according to (5) and (6). This conclusion does not depend on the size of m, nor in particular on whether m remains finite or becomes infinite as $n \to \infty$. It thus applies to ordered factor theories—such as the radex,

with its simplexes and circumplexes (5)—as well as to limited common-factor theories like those of Spearman and Thurstone, whenever the δ-law of deviation (5, p. 308) holds for the unique-factor variables.

Thus, a general sufficient condition for σ_j^2 to tend to u_j^2 when $u_j^2 > 0$ is that $r_j^2 \to 1$ or $r^* \to 1$ as $n \to \infty$. This holds for each j separately.

A less general sufficient condition, and one that does not necessarily hold for any one j but only for "almost all" j, is given in

Theorem 2. If R is nonsingular for all n, and if $\lim_{n\to\infty} m/n = 0$, *then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \frac{u_j^2}{\sigma_j^2} = 1.$$

For all except possibly a zero proportion of the j it must be that $\lim_{n\to\infty} u_j^2/\sigma_j^2 = 1$.

The condition that $m/n \to 0$ holds in particular for the Spearman-Thurstone approach to factor analysis, which postulates that the number of common factors should be small compared to the number of observed variables.

Since $u_j^2/\sigma_j^2 \leqq 1$ for all j, according to (4), we must have the mean ratio also bounded above by unity:

$$\frac{1}{n} \sum_{j=1}^{n} \frac{u_j^2}{\sigma_j^2} \leqq 1. \tag{15}$$

The hypothesis that R is nonsingular for all n ensures that no observed variable is perfectly predictable from the rest, or that $\sigma_j^2 > 0$ for all j and n, so that division by σ_j^2 in (15) is always justified. The first conclusion of Theorem 2 is that the limit of the left member of (15) as $n \to \infty$ is actually the right member. But clearly, the mean value of a sequence cannot tend to an upper bound to each member of the sequence unless almost all members of the sequence also tend to this upper bound. Hence the second conclusion of Theorem 2 follows from the first. We need only to establish the first part of the theorem now.

As is well known, if $R - U^2$ is Gramian and of rank m, we can write

$$R = AA' + U^2, \tag{16}$$

where A is some matrix of order $n \times m$ and of rank m. Let Q be defined as the symmetric matrix of order m:

$$Q = I_m + A'U^{-2}A, \tag{17}$$

where I_m is the unit matrix of order m. It has been shown in (2, 92) that Q is Gramian and nonsingular, and furthermore

$$R^{-1} = U^{-2} - U^{-2}AQ^{-1}A'U^{-2}. \tag{18}$$

It is easily verified further, from (18) and (17), that

$$A'R^{-1}A = I_m - Q^{-1}. \tag{19}$$

Since the left member of (19) is clearly Gramian, so must the right member be. Indeed, it is known that $I_m - Q^{-1}$ is the covariance matrix of the predicted values (from the observed n variables) of any m orthogonal common-factor scores underlying loading matrix A (6). Let q^{kk} denote the kth main diagonal element of Q^{-1}, or the variance of estimate of the kth common factor, and let p_k^2 be defined as

$$p_k^2 = 1 - q^{kk} \qquad (k = 1, 2, \cdots, m). \tag{20}$$

Then p_k^2 is the square of the multiple-correlation coefficient of the kth common factor from the n observed variables, and

$$0 \leqq p_k^2 \leqq 1 \qquad (k = 1, 2, \cdots, m). \tag{21}$$

Therefore, the *trace*—or sum of the main diagonal elements—of $I_m - Q^{-1}$ satisfies

$$\operatorname{tr}(I_m - Q^{-1}) = \sum_{k=1}^{m} p_k^2 \leqq m. \tag{22}$$

We are particularly interested in the trace of U^2R^{-1}, for clearly—remembering (7)—

$$\operatorname{tr}(U^2R^{-1}) = \operatorname{tr}(U^2S^{-2}) = \sum_{j=1}^{n} \frac{u_j^2}{\sigma_j^2}. \tag{23}$$

Since the trace of a product is unchanged if order of multiplication is reversed,

$$\operatorname{tr}(A'R^{-1}A) = \operatorname{tr}(AA'R^{-1}) = \operatorname{tr}(I_n - U^2R^{-1}), \tag{24}$$

the last member following from the middle member by recalling (16). Therefore, taking traces of both members of (19) and using (23), (24), and (22),

$$\sum_{j=1}^{n} \frac{u_j^2}{\sigma_j^2} = n - \sum_{k=1}^{m} p_k^2 \geqq n - m. \tag{25}$$

Dividing (25) through by n and prefixing inequality (15),

$$1 \geqq \frac{1}{n} \sum_{j=1}^{n} \frac{u_j^2}{\sigma_j^2} = 1 - \frac{1}{n} \sum_{k=1}^{m} p_k^2 \geqq 1 - \frac{m}{n}. \tag{26}$$

Clearly, if $m/n \to 0$ in the last member of (26), the middle members must tend to unity, or Theorem 2 is established.

Notice that Theorem 2 could be rephrased to say that almost all $r_j^2 \to 1$, or almost all unique-factors must be determinate in the limit. It is interesting to see this in a slightly different way. From (5) and the middle members of (26),

$$\frac{1}{n} \sum_{j=1}^{n} r_j^2 + \frac{1}{n} \sum_{k=1}^{m} p_k^2 = 1, \tag{27}$$

or

$$\overline{r^2} + \frac{m}{n}\overline{p^2} = 1, \tag{28}$$

where $\overline{r^2}$ is the mean predictability of the n unique-factors, while $\overline{p^2}$ is the mean predictability of the m common factors. When m/n is small, the average predictability of the common factors cannot influence greatly the average predictability of the unique factors: $\overline{r^2}$ must be close to unity. A further consequence is that, if both $\overline{r^2}$ and $\overline{p^2}$ tend to unity as $n \to \infty$, it must be that $m/n \to 0$. This does not require m to remain finite, of course, but only to increase at a less rapid pace than does n.

VI. *Increase in Information with n*

A desirable property of estimates of communalities is that they should improve in general as n increases. Any n variables studied empirically by a factor analysis are usually regarded as but a sample of a far larger universe of variables. The communalities sought are those of the universe.

Of the four approaches to estimates outlined in Part I above, the only one which has its estimates vary explicitly with n is that of lower bounds. In this sense, it is the only one not tied to algebraic artifacts that may arise in data due to the finiteness of n of the observed sample of variables (cf. 3 and 4).

For fixed j, ρ_j^2 must increase with n—or at worst remain constant—for a multiple-correlation cannot become worse as the number of predictors increases. If h_j^2 is defined as for the universe of variables ($n = \infty$), then ρ_j^2 must improve in general as an estimate of h_j^2 as n increases, considering (3). The lower bounds improve as estimates as n increases, taking advantage of the increased information.

Similarly, if the jth unique-factor scores are defined uniquely as for the universe of observed variables, r_j^2 must in general increase with n. From (5), this again makes σ_j^2 an increasingly better estimate of the fixed u_j^2 as n increases.

Thus, the lower bounds automatically take advantage of whatever new information is brought in with increased n, without making any assumptions at all. In broad classes of cases, as we have seen, this new information can make $\rho_j^2 \to h_j^2$ for all or almost all j.

VII. *Further "Best Possible" Inequalities*

We have concentrated until now on the approximation of the σ_j^2 to the u_j^2. Related to this is another problem: the estimation of minimum rank m for Gramian $R - U^2$. We shall show that using the diagonal matrix S^2 of (7) as an estimate of U^2 for finite n leads also to a "best possible" inequality for m, as well as to other important inequalities.

With any nonsingular correlation matrix R is associated another nonsingular correlation matrix R^* defined by

$$R^* = SR^{-1}S. \tag{29}$$

R^* is clearly Gramian, for R^{-1} is Gramian and S is a diagonal matrix. The main diagonal elements of R^* are all unity from the definition of S and the fact that $1/\sigma_j^2$ is the jth diagonal element of R^{-1}. Indeed, R^* is the correlation matrix of the *anti-images* of the n variables of R (cf. **3**, p. 294f). Regardless of the statistical meaning of R^*, it is a perfectly good correlation matrix when n is finite, and we can seek a diagonal Gramian matrix U^* that will leave $R^* - U^{*2}$ Gramian and with minimum rank m^*. This will lead to the interesting and important inequality for the case where no σ_j^2 is a uniqueness nor equals unity:

$$m + m^* \geqq n \qquad (u_j^2 < \sigma_j^2 < 1; j = 1, 2, \cdots, n). \tag{30}$$

The restrictions that $S^2 - U^2$ and $I - S^2$ be nonsingular are essential here (consider the counter-example where $S = R = R^* = I$). That $\sigma_j^2 \neq 1$ ($I - S^2$ be nonsingular) implies that each variable in R has at least one nonzero correlation with some other variable.

According to (30), if m/n is small, then m^*/n must be large. Conversely, if m^*/n is small, m/n must be large. This is rather paradoxical in view of the fact that R^* can *always* be reduced to rank m by subtracting out the diagonal matrix $SU^{-2}S$ ($= S^2U^{-2}$). This follows by pre- and post-multiplying (18) through by S, remembering (29), and noting that the second term on the right is of rank m. Conversely, R can always be reduced to rank m^* by subtracting out $S^{*2}U^{*-2}$, where S^{*2} is the diagonal matrix defined by the main diagonal of R^{*-1}. Thus, if all diagonal-free submatrices of R have rank less than $n/2$, so must those of R^*, and conversely. Regardless, (30) holds.

In effect, then, (30) implies that to every R for which $\sigma_j^2 \neq u_j^2$ or 1 for all j and where $m < n/2$, there corresponds an R^* which is a generalized "Heywood" case (cf. **4**, 159f). Although all diagonal-free matrices have small rank in R^*, no communalities can be found to make $R - U^2$ of equally small rank and yet be Gramian. It must be that $m^* \geqq n - m$. This again emphasizes that the case $m < n/2$ may be the exception, rather than the rule, for correlation matrices. And it is interesting that this paradox arises precisely for those cases where no σ_j^2 equals the corresponding u_j^2.

To establish (30), we first recall the theorem (**4**, 157f) that if $S^2 - U^2$ is nonsingular, and if s is the non-negative index of $R - S^2$, then

$$s \leqq m \qquad (|S^2 - U^2| > 0). \tag{31}$$

Now, the proof of (31) in (**4**) can be modified to take care of the case where $S^2 - U^2$ is possibly singular, to establish the weaker but more universal inequality $p \leqq m$, where p is the *positive* index of $R - S^2$. We shall not take

space to prove this modification here, but shall merely state it in terms of our needs for R^*:

$$p^* \leqq m^*, \tag{32}$$

where p^* is the positive index of $R^* - S^{*2}$, and we do *not* necessarily assume $S^{*2} - U^{*2}$ to be nonsingular.

Now, from (29), $R^{*-1} = S^{-1}RS^{-1}$, or since the main diagonal elements of R are all unity,

$$S^{*2} = S^2. \tag{33}$$

It is interesting to note that (33) and (29) imply that $(R^*)^* = R$, or R is to R^* as R^* is to R.

Statistically, (33) implies that the relative predictability of the jth anti-image from the $n - 1$ remaining anti-images is the same as for the jth original variable from the $n - 1$ remaining original variables. From (29) and (33) we can write the identity

$$R^* - S^{*2} = S(R^{-1} - I)S. \tag{34}$$

Sylvester's "law of inertia" (cf. 4, p. 152) applied to (34) shows that p^* equals the positive index of $R^{-1} - I$, which in turn clearly equals the number of latent roots of R^{-1} greater than unity. Hence p^* equals the number of latent roots of R itself which are *less* than unity. But it has been shown in (4) that s is not less than the number of latent roots of R which are greater than or equal to unity whenever $I - S^2$ is nonsingular. Since R has n latent roots all told, it follows that

$$s + p^* \geqq n \qquad (|I - S^2| > 0). \tag{35}$$

Inequality (30) follows from (31), (32), and (35).

To prove that (30) is a "best possible" inequality, we must show that matrices R exist for which the equality sign holds. It suffices to consider an R which has only two distinct latent roots, say $\lambda_1 > 1$ with multiplicity f and $\lambda_2 < 1$ with multiplicity $f^* = n - f$. Then it must be that

$$m = f, \qquad m^* = f^*. \tag{36}$$

For $m \geqq f$ by inequality (39) of (4, 159), and hence $m = f$ by considering that $R - \lambda_2 I$ is Gramian and of rank f; $m^* = f^*$ by analogous reasoning on R^{-1}. Since $f + f^* = n$, (36) provides a special case where the equality in (30) holds.

Inequality (31) by itself is similarly a "best possible" one. Consider the case where R^* has two distinct latent roots, say $\lambda_1 < 1$ with multiplicity p^* and $\lambda_2 < 1$ with multiplicity $p = n - p^*$. Since $R^{*-1} - I = S^{-1}RS^{-1} - I = S^{-1}(R - S^2)S^{-1}$, p is the positive index of $R - S^2$ while p^* is that of $R^* - S^{*2}$. Also, since no root vanishes, $p = s$ or the positive and non-negative

indices coincide. Since $R^{*-1} - \lambda_2^{-1}I = S^{-1}RS^{-1} - \lambda_2^{-1}I$ is Gramian and of rank $p = s$, so must $R - \lambda_2^{-1}S^2$ be, or the equality in (31) must hold for this case.

VIII. *Relation to Image Analysis*

The ratio of u_i^2 to σ_i^2 indicates the relative predictability of the jth unique-factor scores from the n observed variables of R, according to (5). Closely related is another parameter developed in image theory and denoted by δ_i^2, namely, the variance of the difference between the respective scores on the jth anti-image and the jth unique factor. It turns out (3, 293) that δ_i^2 can be computed as the simple difference

$$\delta_i^2 = \sigma_i^2 - u_i^2 \quad . \quad (j = 1, 2, \cdots, n). \tag{37}$$

Hence, a necessary and sufficient condition that $\sigma_i^2 \to u_i^2$ as $n \to \infty$ is that $\delta_i^2 \to 0$. This implies that the unique-factor scores must be essentially the total anti-image scores from the universe of content. Here we have the individual anti-images themselves as increasingly better estimates of the unique-factor scores as $n \to \infty$. This problem of estimating scores is perhaps even more basic than that of estimating only over-all parameters, such as uniquenesses, which are based on the scores. Estimating U^2 by S^2 has the important property of tying in directly with the score estimation problem via image analysis.

REFERENCES

1. Golub, G. H. On the number of significant factors as determined by the method of maximum likelihood. (Stencilled) Digital Computer Laboratory, Univ. of Illinois, 1954.
2. Guttman, L. Multiple rectilinear prediction and the resolution into components. *Psychometrika*, 1940, 5, 75-99.
3. Guttman, L. Image theory for the structure of quantitative variates. *Psychometrika*, 1953, 18, 277-296.
4. Guttman, L. Some necessary conditions for common-factor analysis. *Psychometrika*, 1954, 19, 149-161.
5. Guttman, L. A new approach to factor analysis: the radex. In Paul F. Lazarsfeld (ed.), Mathematical thinking in the social sciences, Glencoe, Ill.: Free Press, 1954.
6. Guttman, L. The determinacy of factor score matrices, with implications for five other basic problems of common-factor theory. *Brit. J. stat. Psychol*, 1955, 8, 65-81.
7. Rao, C. R. Estimation and tests of significance in factor analysis. *Psychometrika*, 1955, 20, 93-111.
8. Thurstone, L. L. Multiple-factor analysis. Chicago: Univ. Chicago Press, 1947.

Manuscript received 5/17/55

Revised manuscript received 9/13/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

RANK-BISERIAL CORRELATION

EDWARD E. CURETON

UNIVERSITY OF TENNESSEE

A formula is developed for the correlation between a ranking (possibly including ties) and a dichotomy, with limits which are always ± 1. This formula is shown to be equivalent both to Kendall's τ and Spearman's ρ.

Suppose we have two correlated variables, one represented by a ranking (possibly including ties) and the other by a dichotomy. The dichotomy may be considered a ranking concentrated into two multiple ties; its ties, however, do not represent equal measurements (or judgments of equality) on a continuous (or at least a many-step) variable. Rather, the ties represent a broad grouping of the data into two categories, or possibly an actual two-point distribution (sex, e.g.). Since the number of distinct ranks in the ranked variable will always be much greater than 2 and will equal N in the untied case, exact rank agreement of the two variables, pair by pair for each individual, is impossible. In this situation we desire a coefficient which will still have attainable limits ± 1 in all circumstances. It should be $+1$ when all ranks in the "higher" category of the dichotomy exceed all ranks in the "lower" category, and -1 when all ranks in the "lower" category exceed all ranks in the "higher" category. It should be strictly non-parametric, i.e., defined wholly in terms of inversions and agreements between pairs of rank-pairs, without use of such concepts as mean, variance, covariance, or regression. Finally, it should resemble the usual rank correlation coefficients in some reasonable sense.

Let R_x represent the dichotomy, with categories $R_x +$ and $R_x -$, and let R_y represent the ranked variable. Ties in R_y are to be handled by the mid-rank method. We then arrange the ranks R_y in as nearly as possible the natural order $(N, N-1, \cdots, 1)$, with rank N "high" and rank 1 "low," and allocate them to the categories $R_x +$ and $R_x -$ as in the following example:

$R_x +$	$R_x -$	*Inv.*	*Agr.*
9.5			4
9.5			4
8			4
6.5			3
	6.5	2	
4.5			3
4.5			3
	2.5		
	2.5		
	1		
$N_1 = 6$	$N_2 = 4$	$Q = 2$	$P = 21$

(1)

No two R_y ranks may be in the same row, but in case of a tie in R_y with one member falling under $R_x +$ and the other under $R_x -$, the relation between the row and column allocations is immaterial. Thus, in (1), the first 6.5 might as well have been allocated to $R_x -$ and the second to $R_x +$.

With this arrangement, there is an *inversion* at any given number under $R_x -$ for every smaller number under $R_x +$. Thus, at 6.5 in $R_x -$ we have two inversions, one for each of the values 4.5 under $R_x +$. There is also an *agreement* at any given number under $R_x +$ for every smaller number under $R_x -$. Let Q be the total number of inversions, and let P be the total number of agreements.

With this method of allocation to rows and columns, perfect positive correlation would require that all numbers under $R_x +$ should be larger than all numbers under $R_x -$, and in this case we should find that $Q = 0$ and $P = P_{\max}$. Perfect negative correlation would require that all numbers under $R_x +$ should be smaller than all numbers under $R_x -$, and in this case we should find that $P = 0$ and $Q = Q_{\max}$. Also, $P_{\max} = Q_{\max}$, since the two result merely from an interchange of the sets of numbers under $R_x +$ and $R_x -$. Our coefficient may therefore be of the form

$$r_{RB} = (P - Q)/P_{\max} . \tag{2}$$

It will be $+1$ if $Q = 0$ and $P = P_{\max}$, -1 if $P = 0$ and $Q = Q_{\max} = P_{\max}$, and 0 if $P = Q$.

To determine $P_{\max}$, we note first that in the situation in which the coefficient is $+1$, there will be N_2 agreements for every number under $R_x +$, or N_1N_2 in all. There is one case, however, so far passed over, in which $P_{\max}$ cannot be as great as N_1N_2. This case is illustrated in our example. If we set up explicitly the situation for $P = P_{\max}$ with these data, we have:

$R_x +$	$R_x -$	*Inv.*	*Agr.*
9.5			4
9.5			4
8			4
6.5			4
6.5			4
4.5			3
	4.5		
	2.5		
	2.5		
	1		
$N_1 = 6$	$N_2 = 4$	$Q = 0$	$P = 23$
			$N_1N_2 = 24$

(3)

One agreement is lost because the lowest rank under $R_x +$ is tied with the highest under $R_x -$. In other cases there might be a triple or multiple tie at the point of dichotomy. We shall term a tie at this point a *bracket tie.* For any bracket tie, the value of P_{max} will be reduced from N_1N_2 by unity for every *pair* of members of this tie one of which is under $R_x +$ and the other under $R_x -$, after R_y has been rearranged to be as nearly as possible in the natural order and allocation under $R_x +$ and $R_x -$ is made in such a manner as to preserve the original values of N_1 and N_2. If t_1 is the number under $R_x +$ participating in the bracket tie, and t_2 the number under $R_x -$, $P_{max} = N_1N_2 - t_1t_2$, and our formula becomes

$$r_{RB} = \frac{P - Q}{N_1N_2 - t_1t_2}. \tag{4}$$

Physically, it is not necessary to rearrange the original data in order to compute t_1t_2. We merely draw a horizontal line across columns $R_x +$ and $R_x -$ in (1), at a level which leaves N_1 cases above the line and N_2 below it. Since the original arrangement in (1) was with R_y in as nearly as possible the natural order, a bracket tie will then consist of any group of identical numbers, some immediately above and some immediately below this line. The number above is t_1 and the number below is t_2. For the example of (1), we find by (4):

$$r_{RB} = \frac{21 - 2}{(6)(4) - (1)(1)} = .826.$$

Clearly r_{RB} is a Kendall-type coefficient, since Q and P are the numbers of unweighted inversions and agreements, respectively (2). But it is also a Spearman-type coefficient. Durbin and Stuart (1) have shown that, in the untied case, Spearman's coefficient is given by $(U - V)/(U - V)_{max}$, where

V is the number of inversions and U the number of agreements, each weighted by the difference between the two ranks concerned. It is easily shown that the difference which supplies the weight may come from either R_y or R_x, and it is also easy to find $(U - V)_{max}$ for the cases corresponding to Kendall's ρ_a and ρ_b. The writer has not been able to prove in these cases that the values given by $(U - V)/(U - V)_{max}$ are necessarily equal in general to those given by the corresponding formulas based on Σd^2, but he has verified each of them on several sets of numerical data.

In the present case, we need merely note that all R_x values bracketed under $R_x +$ would have one mid-rank value, and all those bracketed under $R_x -$ another. If, then, we weight each inversion and agreement by the corresponding rank-difference in R_x, all weights will be equal (and equal to the difference between the two mid-rank values), and it follows at once that r_{RB} is a Spearman-type coefficient.

The hypothesis that r_{RB} differs only by chance from $\rho_{RB} = 0$ may be tested by the Mann-Whitney extension of the Wilcoxon test (3).

REFERENCES

1. Durbin, J. and Stuart, A. Inversions and rank correlation coefficients. *J. roy. statist. Soc.*, 1951, 13, 303-9.
2. Kendall, M. G. Rank correlation methods. London, Charles Griffin and Co., 1948, Ch. 3.
3. Mann, H. B. and Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, 18, 50-60.

Manuscript received 11/14/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

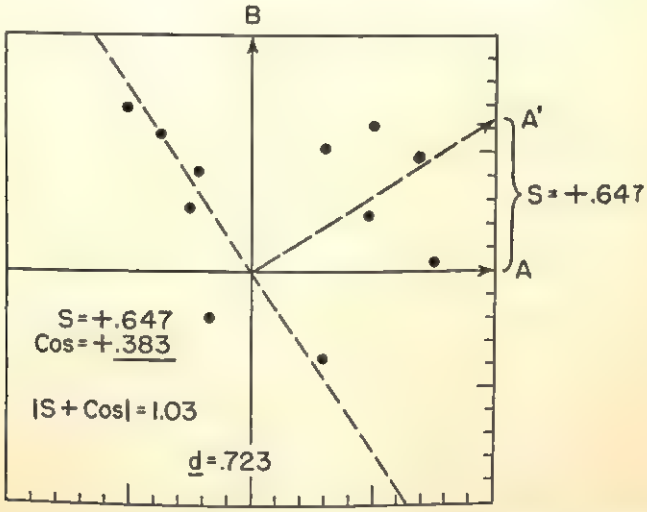
A NOMOGRAM FOR FACTOR ANALYSTS*

P. D. DEUEL

UNIVERSITY OF CALIFORNIA

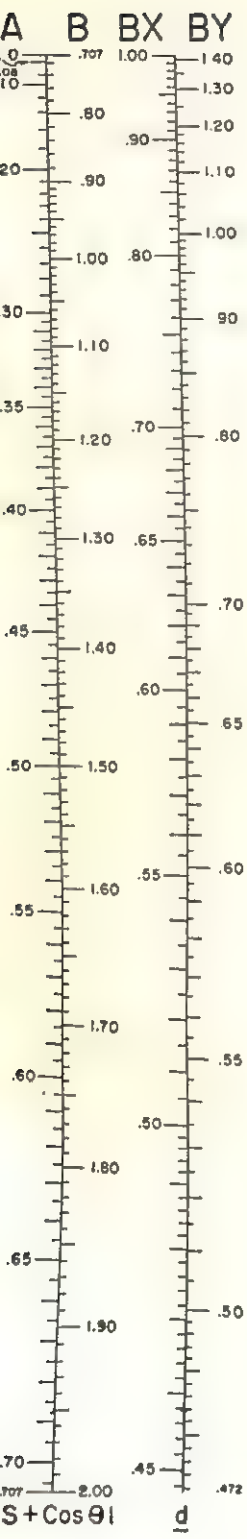
When a new reference vector is chosen graphically from the plane of two old ones, its direction cosines as well as the projections of the tests on it are most easily computed by applying certain multipliers d and Sd to quantities which are already known. The nomogram quickly supplies d, after S has been read from the graph.

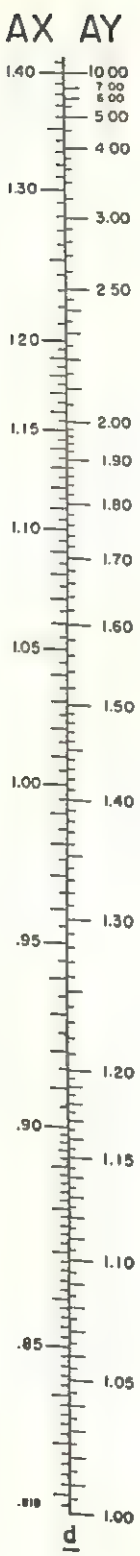
The nomogram accompanying this article reduces the computing work in what is perhaps the most popular of the graphical rotation methods of factor analysis, namely, the diagrammatic method explained in Thurstone (1, pp. 194–216) with rotations made in one plane at a time. Use of the nomogram will be explained in terms of the following figure:



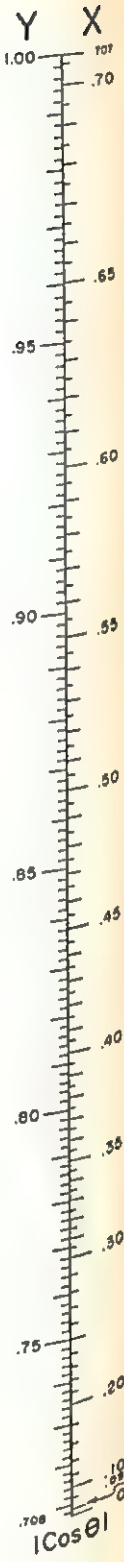
As usual, A and B represent trial reference vectors and the dots represent the projections of the test vector termini in the plane determined by A and B. (Following psychological usage, we refer to the "given" or analyzed variables as tests.) The cosine, .383, of the angle θ between A and B has been recorded in the lower left-hand portion of the diagram; it was obtained

*Suggestions by Norman Livson, Thomas Nichols, and Karyl Atherton have been incorporated in the nomogram. Mrs. Atherton also checked the necessary computations. In addition, I am obligated to Katherine Eardley, scientific illustrator, for her care in lettering and inking the original.





NOMOGRAM giving
$\underline{d} = (S^2 + 2S\cos\theta + 1)^{-1/2}$
as a function of
$|S + \cos\theta|$ and $|\cos\theta|$
for $0 \leq |S + \cos\theta| \leq 2$



in verifying the linear independence of A and B. The figure suggests replacement of A by a new trial reference vector A'', collinear with the dashed-line vector A', for five tests have nearly vanishing projections on A'. From the diagram we read the vector equation $A' = A + .647B$. The projections of all the tests on A'', as well as the direction cosines of A'', are required.

More generally, the method under discussion will always yield a "long reference vector" A' related to known reference vectors A and B by one of the equations $A' = A + SB$ or $-A' = A + SB$, for some number S between -1 and 1. When $A' = A + SB$, the inner product (V, A'') between any vector V and the unit vector A'' collinear with A' is

$$(V, A'') = d(V, A) + S\, d(V, B), \tag{1}$$

where d is $(A', A')^{-\frac{1}{2}}$, the reciprocal of the length of A'. If V is a test vector, (1) gives the desired projection (V, A'') of V on A'' in terms of the known projections (V, A) and (V, B) of V on A and B. Similarly, if V is one of the orthogonal basic vectors, (1) gives the desired direction cosine (V, A'') of A'' with respect to V in terms of the given direction cosines (V, A) and (V, B) of A and B with respect to V. Thus, when d is known, all the required quantities are obtained from (1) by applying the multipliers d and Sd to columns containing the relevant inner products (V, A) and (V, B). The case $-A' = A + SB$ is accommodated by simply negating the right side of (1).

Upon noting that (A, B) is the cosine of the angle θ between A and B, it is seen that

$$d = (|S + \cos\theta|^2 - |\cos\theta|^2 + 1)^{-1/2}, \tag{2}$$

a graphable function of $|S + \cos\theta|$ and $|\cos\theta|$, and in fact, the function represented in the nomogram. To determine d, therefore, it is only necessary to perform the addition $S + \cos\theta$, enter the nomogram with arguments $|S + \cos\theta|$ and $|\cos\theta|$, and read d from the appropriate inside scale. To read the nomogram:

1. Locate $|S + \cos\theta|$ on left-most stem and $|\cos\theta|$ on right-most stem.
2. Align a straightedge through the two points thus found.
3. Pick out appropriate inside scale. This is the scale whose label (at top) combines the labels of the scales on which $|S + \cos\theta|$ and $|\cos\theta|$ were found.
4. Read d where straightedge crosses appropriate inside scale.

Generally, S is read from a graph and cannot be identified with better than two-place accuracy. In such cases a gratuitous third digit may be appended to S, so chosen as to render the third digit of the sum $S + \cos\theta$ zero. This device makes for greater ease and precision in locating values of $|S + \cos\theta|$ near the top of the left-hand stem.

In our example, the value $+.64 \pm$ for S was read from the graph, and the final digit 7 was selected to complement the third digit 3 of $\cos \theta$. The addition $S + \cos \theta = 1.03$ was performed directly on the diagram. The value 1.03 for $| S + \cos \theta |$ was then found on the B-scale of the nomogram; the value .383 for $| \cos \theta |$ was located on the X-scale; hence, $d = .723$ was read from the BX-scale.

The scale factor of the nomogram varies widely from the lower part of the BX-scale to the upper region of the AY-scale. The instrument has been designed, however, with the purpose of securing three-place accuracy for virtually all cases which arise in practice; extensive applications of the nomogram, both to Thurstone's illustrative material and to original factor analyses, indicate that this goal has been well attained. In successive rotations of the above type, the liberal AX-scale carries by far the most traffic.

Larger copies of the nomogram may be obtained by writing to the author at the Institute of Child Welfare, University of California, Berkeley 4, California. The copies have stems approximately $9\frac{1}{4}''$ long and are printed on $8\frac{1}{2}'' \times 11''$ index-card paper.

REFERENCE

1. Thurstone, L. L. Multiple-factor analysis. Chicago: Univ. Chicago Press, 1947.

Manuscript received 8/29/55

Revised manuscript received 10/24/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

A NOTE ON THE ESTIMATION OF NONSPURIOUS CORRELATIONS

WILLIAM H. ANGOFF

EDUCATIONAL TESTING SERVICE

A method is provided for estimating the nonspurious correlation of a part of a test with the total test. Two cases are considered: one in which the actual subtest is parallel to the total test, the other in which the actual subtest is not parallel to the total test.

A problem that frequently arises in the examination of test data is that of estimating the degree of relationship of a test with a subsection of items drawn from the test. In general two methods are available: one is to correlate the subtest with the total parent test; the other is to correlate the subtest with the complementary subtest resulting from the subtraction of the first subtest from the total. In the first method, it is clear that spuriousness exists; even if the subtest is totally unreliable and correlates zero with the complementary subtest, the computed correlation of the part with the total would result in a value greater than zero, and roughly in the proportion represented by the subtest in the total. In general the correlation would be

$$r_{jt} = (\sigma_j + r_{jh}\sigma_h)/\sigma_t, \tag{1}$$

where t is the total test, and j and h are the complementary subsections of the total test. Even if $r_{jh} = 0$, r_{jt} is still greater than zero, merely by virtue of the presence of j in the total test.

The second method, on the other hand, defeats its own purpose—it yields a correlation of the subsection j with the complementary subsection h, not a correlation with a test of the length t.

In order to estimate the nonspurious correlation, say $r_{j't}$, consider test t to be an unspeeded test of power, and, as before, to comprise two parts, j and h. Also consider a hypothetical test, j', exactly parallel and of equivalent effective length (1) to j. Subtests j and h need not be parallel forms. Then

$$r_{j't} = (r_{jj'}\sigma_j + r_{j'h}\sigma_h)/\sigma_t\,. \tag{2}$$

The notation in (2) may be modified slightly. Since j and j' are parallel, $r_{j'h}$ may be written r_{jh}. Also $r_{jj'}$, which expresses the reliability of test j as the correlation between parallel forms, may be written in the conventional notation as r_{jj}. Henceforth, $r_{j'h}$ will be written as r_{jh}, and $r_{jj'}$ will be written as r_{jj}. However, the notation $r_{j't}$, designating nonspurious correlation, will

be retained to distinguish it from the spurious correlation, r_{jt}. With the foregoing modifications in notation the formula for nonspurious correlation is

$$r_{j't} = (r_{jj}\sigma_j + r_{jh}\sigma_h)/\sigma_t \,. \tag{3}$$

In the actual situation, the reliability, r_{jj}, will probably be estimated best by one of the internal consistency reliability formulas appropriate to power tests.

If now a further restriction is placed on test t, namely, that j and h be parallel (but not necessarily of equivalent length), then further simplification is possible. It has been observed (1) that under this restriction

$$r_{jj} = r_{tt}r_{jt}^2 \,, \tag{4}$$

and also

$$r_{tt} = \frac{r_{jt}\sigma_t - \sigma_j}{r_{jt}(\sigma_t - r_{jt}\sigma_j)}. \tag{5}$$

Substituting (5) in (4),

$$r_{jj} = \frac{r_{jt}(r_{jt}\sigma_t - \sigma_j)}{\sigma_t - r_{jt}\sigma_j}. \tag{6}$$

It may also be seen that if $r_{jt}\sigma_t - \sigma_j$ is substituted in (3) for its equivalent, $r_{jh}\sigma_h$ [see equation (1)], then

$$r_{j't} = (r_{jj}\sigma_j + r_{jt}\sigma_t - \sigma_j)/\sigma_t \,. \tag{7}$$

Finally, substituting (6) in (7),

$$r_{j't} = \frac{r_{jt}\sigma_t - \sigma_j}{\sigma_t - r_{jt}\sigma_j}. \tag{8}$$

Two formulas are thus presented for estimating the nonspurious correlation of a subtest of items with the total test from which it is drawn. In (3) and in its equivalent, (7), no restriction is imposed on the kinds of items drawn from the total test. These equations could, for example, represent the estimated correlation of a subset of arithmetic items drawn from a heterogeneous total test consisting of arithmetic, verbal, and spatial items. Equation (8), on the other hand, requires that the subset of items be essentially a short parallel form or miniature of the total test. Whereas (3) and (7) require a separate determination of the reliability of the subtest by means of internal consistency methods, such as the Kuder-Richardson formulas, (8) permits that estimate to be made implicitly as an integral part of the estimate of nonspurious correlation. It is the added restriction that j is parallel to t that makes the simplification possible.

Additional algebraic simplifications may be made in (8). From (1) it is seen that the numerator of (8) may be written $r_{jh}\sigma_h$. It may also be seen

that the denominator of (8) may be written $r_{ht}\sigma_h$ since

$$r_{ht} \equiv r_{(t-j)t} = (\sigma_t - r_{jt}\sigma_j)/\sigma_h \,. \tag{9}$$

Thus (8) may be written

$$r_{j't} = r_{jh}\sigma_h/r_{ht}\sigma_h = r_{jh}/r_{ht} \,. \tag{10}$$

Also, from (6) and (8)

$$r_{j't} = r_{jj}/r_{jt} \,, \tag{11}$$

and finally, from (5) and (8)

$$r_{j't} = r_{tt}r_{jt} \,. \tag{12}$$

REFERENCE

1. Angoff, W. H. Test reliability and effective test length. *Psychometrika*, 1953, 18, 1-14.

Manuscript received 7/28/55

Revised manuscript received 11/7/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

THE VARIANCE OF THE NUMBER OF MUTUAL CHOICES IN SOCIOMETRY

LEO KATZ
MICHIGAN STATE UNIVERSITY
AND
THURLOW R. WILSON
UNIVERSITY OF NEW MEXICO*

The variance of the number of mutual dyads in a sociometric situation where each member of a group chooses independently and at random is derived for unrestricted numbers of choices per group member, as well as for a fixed number of choices. The distribution of the number of mutuals is considered.

I. *Introduction*

In a sociometric test, a number of the pairs of group members (dyads) may show mutuality. A mutual dyad is defined by the two subjects selecting one another. When each member of the group makes a fixed number of choices, the number of mutual dyads has been assumed to have a binomial distribution with p, the probability of a success, equal to the probability that a given dyad is mutual, and n, the number of binomial observations, equal to the number of dyads (1, 7). Several writers have pointed out that since sampling is without replacement, this is not correct (2, 4, 6). Let M represent the number of mutual dyads for a group of N numbers. If each member independently makes d selections at random from the $N - 1$ other members, the probability that a given dyad is mutual is $d^2/(N - 1)^2$, and the number of dyads is $N(N - 1)/2$. The expected value of M is not affected by the non-replacement in sampling and is given by

$$E(M) = np = \frac{Nd^2}{2(N - 1)}. \tag{1}$$

We shall show that for a fixed number of choices

$$\text{Var}\,(M) = E(M)\left(1 - \frac{d}{N - 1}\right)^2, \tag{2}$$

rather than the variance appropriate to the binomial distribution,

$$\text{Var (binomial)} = npq = E(M)\left[1 - \left(\frac{d}{N - 1}\right)^2\right]. \tag{3}$$

*We are indebted to Robert Bush and Hartley Rodgers of Harvard University for helpful criticisms.

The ratio of the binomial estimate to the variance of (2) is $(N - 1 + d)/(N - 1 - d)$; the binomial formula will appreciably overestimate the variance of M if the size of the group is small.

When the members of the group do not make the same number of choices, it has been assumed that one may approximate the variance of M by using the average number of choices, $\bar{d}$, in the formula for fixed d. We shall develop the expression for the variance of M for the general case of unrestricted numbers of choices and examine the approximation with $\bar{d}$. Finally we consider the distribution of the number of mutual dyads.

II. *Variance with Fixed Number of Choices*

We first derive the variance of M for the case where each member makes the same number of choices, d. Define a random variable X_{ij} for the particular dyad composed of individuals i and j so that X_{ij} is 1 or 0 according as dyad ij is or is not mutual. X_{ij} is a binomial variable; thus

$$\operatorname{Var}(X_{ij}) = \frac{d^2}{(N-1)^2}\left(1 - \frac{d^2}{(N-1)^2}\right). \tag{4}$$

Since

$$M = \sum X_{ij}\,, \tag{5}$$

where the sum is taken over all distinct dyads, we have

$$\operatorname{Var}(M) = \sum \operatorname{Var}(X_{ij}) + 2\sum \operatorname{Cov}(X_{ij}X_{kl}). \tag{6}$$

The covariances are summed for all distinct pairs of dyads.

If two dyads are composed of four different individuals, the occurrence of mutual choice for the first dyad is independent of the occurrence of mutual choice for the second; hence the covariance of such a pair is zero. A pair of dyads may not have two members in common, but a pair may have one person in common. There are $N(N-1)(N-2)/2$ such pairs. The expected value of the product $X_{ij}\cdot X_{ik}$ is just the probability that both are equal to 1:

$$E(X_{ij}\cdot X_{ik}) = \frac{d^3(d-1)}{(N-1)^3(N-2)}\,, \tag{7}$$

where dyads ij and ik have one member in common. We find the covariance for an overlapping pair of dyads:

$$\operatorname{Cov}(X_{ij}X_{ik}) = -\frac{d^3(N-1-d)}{(N-1)^4(N-2)}. \tag{8}$$

Substituting the values from (4) and (8) in (6) and simplifying we obtain the variance of mutual choices given by (2).

III. *Variance with Unrestricted Numbers of Choices*

We now derive the variance for the case where the members make unrestricted numbers of choices. Let d_i denote the number of choices actually

made by subject i. The probability that dyad ij is mutual is given by

$$p_{ij} = \Pr(X_{ij} = 1) = \frac{d_i\, d_j}{(N-1)^2}. \tag{9}$$

The probability that both dyad ij and kl are mutual is

$$p_{ij,k} = \Pr(X_{ij} \cdot X_{kl} = 1) = \frac{d_i\, d_j\, d_k\, d_l}{(N-1)^4}, \tag{10a}$$

if all four individuals are different, while

$$p_{ij,ik} = \Pr(X_{ij} \cdot X_{ik} = 1) = \frac{d_i(d_i - 1)\, d_j\, d_k}{(N-1)^3(N-2)}, \tag{10b}$$

if the two dyads consist of three different members.

Conventional methods for obtaining the variance (5, pp. 60 ff.) involve computing two sums,

$$S_1 = \sum p_{ij}, \tag{11a}$$

$$S_2 = \sum p_{ij,kl}, \tag{11b}$$

where both sums are evaluated over all distinct sets of subscripts. Considering (5), (9), and (11a) we see that the mean of the number of mutuals is S_1.

To express Var (M) as a function of S_1 and S_2 we first note that

$$\operatorname{Var}(X_{ij}) = p_{ij} - (p_{ij})^2, \tag{12}$$

and

$$\operatorname{Cov}(X_{ij}X_{kl}) = p_{ij,kl} - p_{ij}p_{kl}. \tag{13}$$

Summing over the variances and covariances yields

$$\sum \operatorname{Var}(X_{ij}) = S_1 - \sum (p_{ij})^2, \tag{14}$$

and

$$\sum \operatorname{Cov}(X_{ij}X_{kl}) = S_2 - \sum p_{ij}p_{kl}. \tag{15}$$

Since

$$S_1^2 = \sum (p_{ij})^2 + 2 \sum p_{ij}p_{kl}, \tag{16}$$

substitution of (14) and (15) in (6) produces, after simplification,

$$\operatorname{Var}(M) = S_1 + 2S_2 - S_1^2. \tag{17}$$

We turn next to the computation of S_1 and S_2. The value of S_1 is by definition

$$S_1 = \frac{1}{(N-1)^2} \sum_{i<j} d_i\, d_j. \tag{18}$$

It is, of course, necessary to require $i < j$ in order to prevent duplication of cases in the sum. The summation appearing in (18) is the second elementary sum of the numbers d_i, usually written

$$a_2 = \sum_{i<j} d_i\, d_j\ . \tag{19}$$

These are related to the more familiar power sums,

$$s_m = \sum d_i^m\ , \tag{20}$$

by the relations (3)

$$a_2 = \frac{1}{2}(s_1^2 - s_2), \tag{21a}$$

$$a_3 = \frac{1}{3!}(s_1^3 - 3s_2s_1 + 2s_3), \tag{21b}$$

$$a_4 = \frac{1}{4!}(s_1^4 - 6s_2s_1^2 + 3s_2^2 + 8s_3s_1 - 6s_4). \tag{21c}$$

Those shown will suffice for the present computations. In particular, we have established from (18), (19), (20), and (21a) that the mean number of mutuals is

$$E(M) = \frac{1}{2(N-1)^2}[(\sum d_i)^2 - \sum d_i^2]. \tag{22}$$

If all d_i equal d we obtain (1) above.

The value of S_2 is somewhat more involved. In the first place, each set of four different persons may form mutual pairs in three ways. Secondly, each set of three, distinct, may have any of the three at the center of the chain of two mutual choices. Taking both of these features into account, we have

$$\begin{aligned} S_2 = {} & \frac{3}{(N-1)^4} \sum_{i<j<k<l} d_i\, d_j\, d_k\, d_l \\ & + \frac{1}{(N-1)^3(N-2)} \sum_i \{ \sum_{\substack{j<k \\ \neq i}} d_i(d_i - 1)\, d_j\, d_k \}. \end{aligned} \tag{23}$$

Making use of (21a) and (21b) we have, after some reduction,

$$\begin{aligned} S_2 = {} & \frac{1}{8(N-1)^4}(s_1^4 - 6s_1^2s_2 + 3s_2^2 + 8s_1s_3 - 6s_4) \\ & + \frac{1}{2(N-1)^3(N-2)}(s_1^2s_2 - s_2^2 - 2s_1s_3 + 2s_4 - s_1^3 + 3s_1s_2 - 2s_3). \end{aligned} \tag{24}$$

Combining this with previous computations, we obtain

$$\begin{aligned}\operatorname{Var}(M) = {} & \frac{1}{2(N-1)^4}(-2s_1^2s_2 + s_2^2 + 4s_1s_3 - 3s_4) \\ & + \frac{1}{(N-1)^3(N-2)}(s_1^2s_2 - s_2^2 - 2s_1s_3 + 2s_4 - s_1^3 + 3s_1s_2 - 2s_3) \\ & + \frac{1}{2(N-1)^2}(s_1^2 - s_2).\end{aligned} \tag{25}$$

When all d_i equal d, (25) reduces to (2).

As examples with unequal numbers of choices, consider two sociometric measurements on 10 individuals. In the first case, two subjects choose two persons each, five choose three, and three choose four. The power sums are 31, 101, 343, and 1205 for s_1, s_2, s_3, and s_4, respectively. Equations (22) and (25) give $E(M)$ equal to 5.31 and Var (M) equal to 2.30. Observe that the use of $\bar{d}$ equal to 3.1 in (1) and (2) would give the very close approximations, $E(M)$ equal to 5.34 and Var (M) equal to 2.29. This happy situation would not obtain if the numbers of choices were considerably more variable, as in the second example; five make a single choice and five make seven choices. The power sums are 40, 250, 1720, and 12,010. From (22) and (25) $E(M)$ is 8.33 and Var (M) is 2.33. With $\bar{d}$ of 4, we obtain the values 8.89 for $E(M)$ and 2.74 for Var (M).

IV. *Distribution of the Number of Mutual Dyads.*

In principle, we could determine the distribution of M exactly by conventional methods (5, p. 64). Denote the maximum possible number of mutual dyads for a group by Max. We define S_m as suggested by (11a) and (11b) above. Then

$$\begin{aligned}\Pr(M = m) = S_m - \binom{m+1}{m}S_{m+1} + \binom{m+2}{m}S_{m+2} \\ - \cdots - + \binom{\text{Max}}{m}S_{\text{Max}}.\end{aligned} \tag{26}$$

We have used (26) only for the case where each person makes a single choice, where

$$S_m = \frac{\binom{N}{2m}\binom{2m}{2}\binom{2m-2}{2}\cdots\binom{2}{2}}{m!(N-1)^{2m}}, \tag{27}$$

and Max equals $N/2$.

We conjecture that for large groups with roughly equal d, M has an Approximately Poisson distribution; from (2) one readily notes that with

increasing group size and r held constant, the variance of M approaches $E(M)$ and the covariance term approaches zero.

The senior author is currently engaged in determining higher moments for various common fixed d by methods similar to those used to produce the variance for the general case. These higher moments will permit determination of the distribution of M for small groups.

REFERENCES

1. Bronfenbrenner, U. A constant frame of reference for sociometric research. *Sociometry*, 1943, **6**, 363–397.
2. Crisswell, Joan H. Foundations of sociometric measurement. *Sociometry*, 1946, **9**, 7–13.
3. David, F. N. and Kendall, M. G. Tables of symmetric functions—part I. *Biometrika*, 1949, **36**, 431–449.
4. Edwards, Daisy S. The constant frame of reference problem in sociometry. *Sociometry*, 1948, **11**, 372–379.
5. Feller, W. Probability theory and its applications. Vol. 1, New York: Wiley, 1950.
6. Loomis, C. P. and Pepinsky, H. B. Sociometry, 1937–1947: theory and methods. *Sociometry*, 1948, **11**, 262–286.
7. Moreno, J. L. and Jennings, Helen H. Sociometric measurement of social configurations, based on deviations from chance. *Sociometry Monographs*, No. 3, 1945, Beacon House.

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

A NOTE ON JENKINS' "IMPROVED METHOD FOR TETRACHORIC r"

JOSHUA A. FISHMAN

COLLEGE ENTRANCE EXAMINATION BOARD

Some readers who will be delighted to utilize Jenkins' method and tables for estimating the tetrachoric correlation (1) may be puzzled to discover that no explicit provision for negative correlations is included. If we follow Jenkins' instruction to "letter the fourfold table so that a is smaller than d and ad greater than bc" four possible arrangements may obtain:

$$1.\ \begin{array}{c|c} c & d \\ \hline a & b \end{array} \qquad 2.\ \begin{array}{c|c} b & a \\ \hline d & c \end{array} \qquad 3.\ \begin{array}{c|c} a & b \\ \hline c & d \end{array} \qquad 4.\ \begin{array}{c|c} d & c \\ \hline b & a \end{array}$$

Of the above four arrangements, the first (which is the one illustrated by Jenkins) and the second are indicative of a positive correlation and the third and fourth are indicative of a negative correlation. If either of the latter two arrangements does obtain, then the final correction [obtained by multiplying the base correction by the multiplier, as in step 5, p. 257, (1)] should be (algebraically) *added to*, rather than subtracted from the negative uncorrected r to obtain the corrected tetrachoric r, which should, of course, be given a negative sign. The important fact to keep in mind is that the correction always reduces the *absolute* size of the uncorrected tetrachoric r.

A few words also might be in order concerning the location of decimal points in Tables 2 (Base Correction) and 3 (Multipliers for Base Correction). Whereas in the former table the omitted decimal points consistently belong before the first digit of the reported three-digit table entries, in the latter table the omitted decimal points belong before the first digit of two-digit table entries, between the first and second digit of three-digit table entries, and followed by a zero for one-digit table entries. Thus, whereas a table entry of 106 should be understood as .106 in Table 2, it should be taken as 1.06 in Table 3. Furthermore, 90 is .90 (as illustrated by Jenkins, p. 257) but 9 is .09 in Table 3.

1. Jenkins, W. L. An improved method for tetachoric r. *Psychometrika*, 1955, 20, 253-258.

BOOK REVIEW

Mathematical Models of Human Behavior. Proceedings of a Symposium Sponsored by Dunlap and Associates, Inc., and the Commission on Accidental Trauma, Armed Forces Epidemiological Board. 1955. vii + 103 pp.

Those who expect to read this book in its entirety, or nearly so, would do well to turn first to Professor Lazarsfeld's Concluding Remarks, since they provide some degree of unification for what is otherwise a rather disjointed collection of reports on several diverse lines of investigation. Those who do not expect to read the entire book would do well to read at least Professor Lazarsfeld's Concluding Remarks for a brief but lucid statement of the function of models, especially in the behavioral sciences. He distinguishes between static and dynamic models, makes passing reference to the predictive function of models, and concentrates on their linguistic function. The linguistic function he divides into three parts: organizing, analytical, and mediating. Naturally his remarks refer to the papers in the Symposium, but they are quite intelligible in themselves.

The Symposium itself was held in February of 1954, in connection with a study being made by Dunlap and Associates, Inc., on the application of mathematical techniques to the study of accidents. Since accidents are "partly the result of human behavior," and since many experts were already engaged in devising and studying mathematical models of human behavior, the proposal was made to invite some of these experts to meet together to describe their work and to participate in informal discussions. This publication contains the papers, but unfortunately not the discussion.

Of the ten papers included (not counting the Concluding Remarks, and not counting a paper by Lorge and Solomon to be published elsewhere) only one deals with accidents. This is one by H. H. Jacobs, who discusses the difficulties in trying to separate the effects of contagion from individual differences as to liability. Two of the speakers, Bush and Estes, discussed somewhat related stochastic learning models. The other seven papers had to do, more or less directly, with utility, or decision making, or both.

For general background on models, the paper by Coombs and Kao might be classed along with Lazarsfeld's Concluding Remarks. This is actually the first section of a report on multidimensional analysis, and one feels suspended in mid-air at the end. As an inducement to learn more, this paper succeeds very well. At the other extreme, a paper by Luce on the formation of coalitions in game theory is largely for the experts. The other five papers are more self-contained, and more directly concerned with utility and decision as such.

Professor Lazarsfeld gently chides the speakers for their preoccupation with gambling, although, as one of them remarks, the gambling situation provides the most direct and realistic contact with the individual's utility function. Merrill Flood's little "Group Preference Experiment" does not make use of a conventional gambling situation, but it does afford, to each of the subjects, a possibility, without certainty, of some gain. A collection of objects is shown to a group of individuals, and certain broad conditions are laid down according to which one of these objects can be had for the group to dispose of. They are left to decide which object it shall be and how it shall be disposed of. Disposition might be by lot to one of the group, or by auction or sale with proceeds divided, and other possibilities can be conceived.

Marschak is concerned with decisions by individuals, pointing out that even when the outcome of an act is known with certainty the same individual may make different choices at different times. He considers various hypotheses that might be made in formulating a model of such inconsistent behavior. Markowitz discusses a hypothesis of Friedman and

Savage for explaining insurance and lotteries; Markowitz replaces it by one of his own which seems to accord better with well-known facts. Jarvik's discussion of gambling is largely discursive, and Edwards describes a series of experiments on gambling which endeavor to arrive at utilities.

The success of the Symposium as such could be judged best from the discussion which is not published. As a publication, it is interesting for showing the diversity of activities under way, but unsatisfying just because of the diversity. As an issue of a periodical, this little volume would do very well. It lacks the cohesion to stand well by itself.

The proofreading is rather poor; presumably the spelling "baracentric" on page 22 is a typographical error.

A. S. Householder

PSYCHOMETRIKA—VOL. 21, NO. 4
DECEMBER, 1956

A TRIBUTE TO L. L. THURSTONE*

Louis Leon Thurstone's contributions to the development of psychology as a quantitative rational science are among the major scientific achievements of the present century. Many of us here still remember the thrill of learning about his early developments in psychophysical scaling methods during the 1920's. The psychophysical measurement methods developed in the latter part of the nineteenth century were used to measure the functional relation between the physical intensity of a stimulus, such as a light or a sound, and the psychological intensity of the sensation—the brightness of the light, the loudness of the sound. Thurstone was among the first to point out that these methods could be modified to construct a scale for measuring psychological qualities that had no measurable physical correlate. He developed the law of comparative judgment and demonstrated that in conjunction with the method of paired comparisons it could be used to measure purely subjective attributes, such as the aesthetic merit of paintings or the strength of an attitude.

In all his work, he stressed the fact that as long as we have merely a rule of procedure for analyzing data, we have no science. He insisted that every theory must be so precisely stated that one of the possible conclusions would be that the data collected were in disagreement with the theory.

In particular for the law of comparative judgment, he developed the criterion of internal consistency for psychophysical scales. For linear scales this means that the distance from object i to object j (as determined from one set of judgments "i greater than j") plus the distance from j to k (as determined from *another* set of judgments "j greater than k") should give a sum in reasonable agreement with the distance from i to k (as independently given by judgments "i greater than k"). Such a statement must be true for all possible sets of three, so that if one has as many as ten or twenty objects in the scaling experiment the number of checks becomes very large.

This, of course, is merely one illustration of the application of the criterion of internal consistency. One of Thurstone's important contributions was his insistence that each experimental and analytical procedure must contain such internal checks.

In addition to devising the theoretical framework for such psychological measurement, we remember Professor Thurstone for numerous applications of these methods to practical problems that were carried out by him and

*A statement read by Harold Gulliksen on behalf of the Psychometric Society at the annual meeting of the Society, September 4, 1956.

his students. One of the early applications was to measurement of the effects of various movies on attitudes of children; other studies measured the effect of various propaganda devices in changing attitudes, or measured the changes in national opinions as reflected in newspapers over a period of years. This area of psychological scaling opened by Thurstone over thirty years ago is still developing both in its theoretical aspects and its uses. Applications of these methods to measurement of intensity of attitudes and to the precise comparison of value systems have been or are being made. The psychophysical scaling methods are important and powerful scientific tools. During the coming decades, prolific and fruitful use of these methods in the development of such fields as linguistics, sociology, cultural anthropology, political science and economics will probably be seen.

Thurstone's achievements in psychology cannot be properly appreciated unless seen against the backdrop of psychological developments of the last half-century. Today it is taken for granted that aptitude and achievement tests can predict various types of academic and job performance with a useful degree of accuracy. Throughout this country, some hundreds of thousands of persons take some millions of tests annually. How often do we stop to remember that prior to 1900 there were literally no aptitude tests available for prediction of academic or job performance? Prediction of educability of children was one of the critical unsolved problems of democratic society. Commissions were appointed in various countries to study the problem. The Binet test was developed in France and was useful in predicting school achievement but was an unwieldy and cumbersome instrument. During the First World War, the Army Alpha and Army Beta tests were developed and used in the first large-scale mass testing program in the history of the world. L. L. Thurstone, as a young psychologist, was active in this testing program.

Thurstone saw the inadequacies in the then widely accepted notions of "general" intelligence. He realized that new and more powerful methods must be developed to analyze the masses of data necessary for a thorough study of the different aspects of cognitive ability, or as we now say, the different factors, or primary mental abilities. He made a very simple change in the then current theory. Instead of assuming that each person was to be characterized by *one* number G (his general intelligence), Thurstone assumed that it might take a great many numbers to describe the person—one number for each of the primary mental abilities.

He outlined his development of this problem to mathematicians whom he knew and thus learned that a field of mathematics called "matrix theory" existed. At over forty years of age, he decided to master a new field of mathematics, because it might help him in analyzing the nature of human abilities. He tutored regularly, worked problems, studied different texts—the result was the development of the factor methods that have been applied exten-

sively by Thurstone, his students, and others, not only in analyzing the domain of mental ability but also in studying blocs in a legislature, schools of thought among teachers regarding curriculum content, classifying allergies, analyzing anthropometric measurements, organizing psychotic symptoms, and so on. The factor methods are extremely powerful, and opportunities for their application in the study of human behavior seem to be limitless. These developments and applications stem in large measure from Thurstone's gift for seeing an important problem, defining it clearly, and then sparing no effort in his persistent search for a solution.

This development of testing over the past fifty years should be seen not only as a scientific achievement but also as a humanitarian accomplishment. Fifty years ago any young person who wished to enter on a given course of study had no alternative but to try, if he were permitted to, for months or even for years; eventually by trial and error he would succeed or be discouraged by repeated failures and cease trying. Now the inept student need not face the discouragement of tackling too high a goal, and the gifted student need no longer be dependent on attracting the attention of some influential person in order to obtain opportunity for advanced training. The unknown person of talent can be identified and encouraged to proceed with advanced training in some appropriate area. Such identification and utilization of various degrees of talent can be of enormous benefit to society and also a boon to the individual who derives added happiness from engaging in an occupation which matches his abilities.

Fifty years ago no aptitude tests were available for such prediction of various aspects of human performance. That we can now see our way reasonably clearly to this goal is in no small measure due to the theoretical and the practical contributions of Professor Thurstone to the field of psychological measurement. His outstanding contributions to the fields of factor analysis and psychophysical scaling by no means exhaust the range of his achievements toward developing psychology as a quantitative rational science. He worked on developing personality measures, and on a quantitative rational learning theory. His Ph.D. thesis of 1916 constitutes one of the early attempts to develop equations of the learning curve.

On his retirement from the University of Chicago, his work continued to be as productive and ingenious as ever. At the University of North Carolina he was leading projects concerned with developing new psychophysical scaling methods and developing a set of novel ideas for obtaining objective measurements of personality characteristics. It is to be hoped that these ideas will not be lost, but that some research worker or group of research workers will develop and validate personality tests along the lines indicated by Thurstone.

During more than a quarter century at the University of Chicago, in addition to making outstanding contributions in many scientific areas Pro-

fessor Thurstone trained large numbers of students who came not only from the United States but also from various foreign countries to study with him—to learn the techniques he was developing, the general principles of scientific investigation, and the principles of the quantitative rational approach which he espoused.

Louis Leon Thurstone has a unique position among psychologists of this century as an original research worker and as an inspiring teacher. Many of us who were privileged to know him closely found him a helpful and understanding friend.

We, the members of the Psychometric Society, feel it our particular privilege to pay homage to his memory. He conceived of this Society and its journal, *Psychometrika*. To his efforts more than to those of any other individual, both the Society and *Psychometrika* owe their present status and even their very existence.

As psychologists, we feel that our past achievements and our future aspirations in the theory and practice of psychometrics have been greatly influenced by Thurstone's developments, by his insights, and by his standards. We feel ourselves pledged to further his ideal: the development of psychology as a quantitative rational science. The greatest honor we can accord him as an outstanding scientist lies in our resolve to continue the development of psychology in the rigorous tradition he did so much to establish.

PSYCHOMETRIKA—VOL. 21, NO. 4
DECEMBER, 1956

THE SCALE GRID: SOME INTERRELATIONS OF DATA MODELS*

CLYDE H. COOMBS†

UNIVERSITY OF MICHIGAN

Perhaps an appropriate subtitle for this paper would be "Some Speculations on the Interrelations of Psychological Methodologies." The *Scale Grid* is a name I have given to a model which presumes to define the underlying continuities between such diverse areas as psychophysics, objective testing, attitude studies including questionnaire and interview techniques, learning experiments, rating scale methods, essay examinations, and projective instruments. The intent of the Scale Grid is to make explicit the fundamental similarities and differences of the methodologies among these various areas of psychological research. The increasing abundance of models and methods in all of these areas, with their associated nomenclature and specialized vocabularies, makes a unification of them increasingly desirable.

In some of these areas of psychological research, serious and intensive efforts have been made to construct models on a genotypic level to explain and predict manifest behavior. One thinks here, for example, of the area of signal detection in psychophysics and of objective test performance. In other areas the models are less explicit and tend to be on a literary level. However, even in these latter areas one may look at the methods of analyzing data. Because there is always a model, at least implied, some of the elements of the models are evident. One may look at this universe of models, explicit and implicit, abstract certain universal elements, and try to characterize them.

Our starting point will be to determine the primitive datum in psychology. What we want is an abstract definition which will hold for every type of psychological observation. Let us begin by taking some examples, seeing what the basic abstract datum is in each case, and then formulating a general definition. In some psychophysical experiments, for example, individuals judge which one of several stimuli is the greatest. In view of what the experimenter subsequently does with the data, it is evident that he thinks of each stimulus as a point; the judgment of the individual is inter-

*Presidential Address to the Psychometric Society, September 4, 1956.

†I wish to express my deep appreciation of the courtesies, assistance, and critical audience furnished me in the preparation of this paper by Professor H. C. J. Duijker and Professor A. DeGroot of the University of Amsterdam, while I was a Fulbright Research Fellow there in 1955-56.

preted as an order relation on these points. Over many replications, the model may deal with a distribution of points for each stimulus and be a probability or actuarial model, but this is not the type of distinction which is at all important to us now. Whether the models are deterministic or actuarial is not relevant to the fundamental distinctions between methodologies I want to make. In fact, I want only to make the point now that in some types of experiments the manifest behavior is interpreted as an order relation between a pair of points, both of which are identified with stimuli.

A different case arises, however, when an individual takes a mental test and passes some items and fails others. The use to which the behavior is put also suggests that it is being interpreted as a relation on a pair of points, but here one point is identified with an item and the other point with the individual. The point associated with the individual represents a measure of his ability; the point associated with the item represents its difficulty. The behavior of the individual in passing or failing the item is interpreted as an order relation on this pair of points. I am not here concerned with the numerical scores on a test or even how a theory arrives at such a score from the basic datum. These are differences on a higher level, and I am here concerned only with differences on most the primitive quantitative level.

When an individual is given an attitude scale and asked to indicate which items he will indorse, again the behavior is interpreted as a relation on a pair of points—one a stimulus, the other an individual. Here the relation is on a psychological distance between the point associated with an individual and the point associated with the stimulus. If the point associated with the stimulus is "near," in a sense defined by the model, the point associated with the individual, he indorses the item, otherwise not. So the behavior is interpreted as indicating whether the distance between two points is greater or less than a certain amount.

When an individual is asked to place a stimulus on rating scale, again the behavior is interpreted as a distance between the point associated with the stimulus and a point associated with a response category which is just another stimulus. Consider, for example, an individual who is asked to rate a stimulus, say a picture, as to whether it is superior, good, or poor. The picture is conceived of as being a point on the scale for this individual; the three points on the rating scale, superior, good, and poor, are also three stimulus points. From the latter points the individual selects the one nearest the point corresponding to the picture. So we see that rating scale behavior is interpreted as a relation between points. The same analysis holds if the rating scale is an ordered set of numbers or the real line. In fact counting is also rating scale behavior. If we ask an individual how many students there are in a class it does not matter whether he guesses or counts as far as the basic datum is concerned. The response, e.g., "35," is interpreted as

a relation between one stimulus, the size of the class as perceived, and another stimulus, a real number.

When an individual is asked whether he observed a light increment or not, the behavior is interpreted as a relation between a point identified with the individual, a threshold, and a point identified with the stimulus, the magnitude of the increment. For a final example, consider an individual asked to judge which of two pairs of stimuli is more similar. Here the behavior is usually interpreted as a relation between two distances; if each distance is interpreted as a point, then behavior implies a relation on a pair of points.

All of these examples illustrate one important fact: behavior is made into data by interpreting it as an order relation between points or a relation on distances—both may more generally be regarded as a relation on a pair of points. An important distinction must be drawn between behavior and data. A datum is defined in this paper as a relation between points. That this is not a new idea is evident from a half-page note by Madison Bentley (1) in which he speaks of Stumpf, Wundt, Ebinghaus, Mach, G. E. Müller, and others who took the view that psychological measurement is a distance measurement, which is just a special case of a relation between pairs.

We have been speaking here as if there were just a single distinction between behavior and data. Actually a threefold distinction should properly be made. We may use the term *behavior* to refer to anything observable about the individual, *raw data* to refer to that which is selected for analysis, and *data* to refer to the interpretation of the *raw data* as an abstract relation between points. The first step in going from behavior to raw data, deciding what to observe, is a many-faceted problem which lies outside the scope of this paper. We are here concerned exclusively with the raw data and how it is interpreted as data in the sense defined above. I shall pursue this distinction between the *raw data* and the *data*, illustrating it in detail shortly.

In principle one could put any data in a matrix as follows: If the data were a relation on a pair of stimuli then a square matrix with rows and corresponding columns identified with stimuli (cf. Figure 1) could nicely accommodate the data. Each cell would contain an entry indicating the relation between that corresponding pair of stimuli. Another experiment, where one member of the pair of points was identified with an individual and the other with a stimulus would require that the matrix of Figure 1 be expanded as in Figure 2. Thus an experiment in which one point of a pair of points was identified as a stimulus and the other as an individual would be entered in the left portion; another experiment in which the behavior was interpreted as a relation on a pair of points, both of which were identified as stimuli, would be entered in the right portion.

Now if we go a little further and consider an experiment where the members of the pair of points are both identified with individuals, the

matrix of Figure 2 becomes as in Figure 3. According to this figure, the behavior observed in some experiments is interpreted as a relation on pairs of points in which both points may be identified with stimuli, both points may be identified with individuals, or one point with a stimulus and one point with an individual. When the data are relations between stimulus

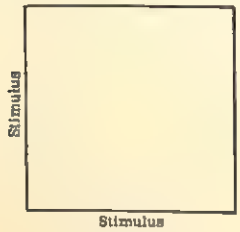


FIGURE 1

A Data Matrix

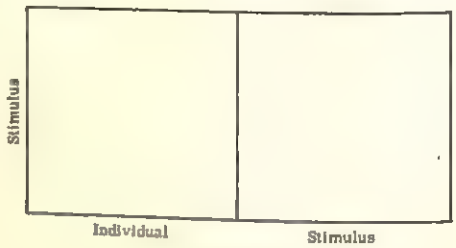


FIGURE 2

A Data Matrix

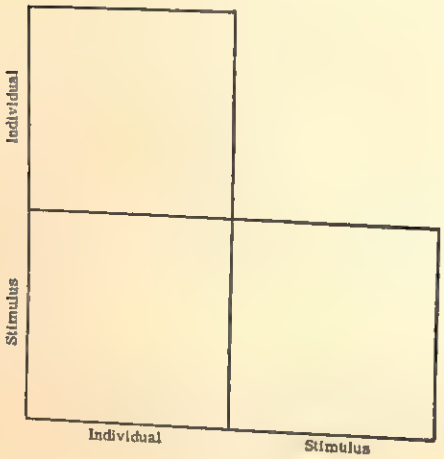


FIGURE 3

A Data Matrix

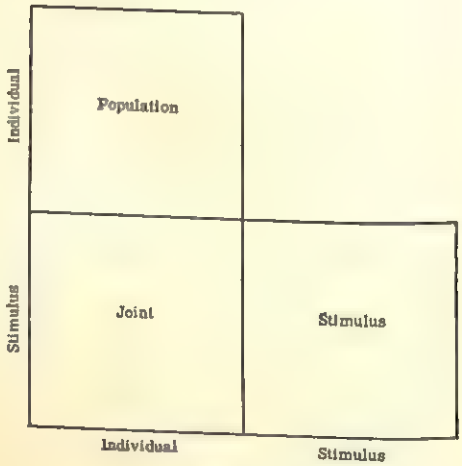


FIGURE 4

A Data Matrix

points, the analysis yields the order of these points on a psychological attribute or the location of these points in a multidimensional space; the individuals who made the judgments or responses are not located as points in the space. So we might call such a space a *Stimulus* space. Correspondingly, we can talk about a space in which only individuals are located as a *Population* space, and a space with both stimuli and people as a *Joint* space. We have the kinds of spaces in which psychological data are analyzed classified in Figure 4.

We now have the beginning of what I call the *Scale Grid*. We could move in either of two directions: developing the model and putting a little more meat on the skeleton, or constructing a psychological interpreta-

tion in order to bring out the implications of the grid. I find it rather difficult not to do both as they are so closely interdependent. So I shall first develop some of the theoretical ideas which will be most related to the interpretations which will follow. We shall consider some typical experiments which are mapped into Joint spaces and some which are mapped into Stimulus spaces; then we shall see what the difference is between them. The characterization of this difference will constitute one dimension of the Scale Grid.

Model Underlying The Scale Grid.

Behavior on a mental test is interpreted as a relation on a pair of points, one of which is associated with the individual and the other associated with the item or stimulus. Such behavior is mapped into data which, when analyzed, yields a Joint space with both stimuli and individuals located in it. What is the primitive operation here? The test item was conceived of as having a certain difficulty and the individual was, in effect, asked to compare his ability level with the difficulty level of the item. In a psychophysical study of the thresholds of individuals, the same is true, e.g., the individual is asked whether he perceives the stimulus, *yes* or *no*, and the behavior is interpreted as a relation on a pair of points, one of which is associated with the threshold of the individual, the other with a stimulus magnitude.

On the other hand, in a psychological study designed to measure heaviness of weights, length of lines, brightness of lights, or what have you, what is the primitive operation? The stimuli are conceived of as points on an attribute continuum, and the individual is asked which of the stimuli is greater. The behavior is interpreted as data on pairs of points both of which are associated with stimuli. Analysis of the data locates the stimuli on a scale, but no attempt is made to locate the individual as a point on the scale. The result is a Stimulus space.

Suppose I have some attitude statements about the church. I want to scale the items and then measure people's attitudes with them. The first step is to scale the items, so we ask individuals to evaluate the items as to which is more pro-church. A data matrix is constructed which is analyzed to yield a Stimulus space, say a one-dimensional scale, with the items located as points on a continuum. We note that the individuals were asked to evaluate the items with respect to where the items were on the continuum and not with respect to any *point* on this continuum which corresponded to the individual. Having scaled the items, we turn around and ask the individual which items he indorses. When we analyze these data, we end up with the individuals located on the same continuum, because this time the experimenter gave the individuals a different task to perform. The individuals were each asked to evaluate the items with respect to some point on the continuum corresponding to his own attitude toward the church, so now we are in a Joint space.

I could go on with examples from conditioning experiments or studies in perception, etc., but will not take the time for it. The important thing is to see the essential difference between behavior which is mapped into a Joint space and behavior which is mapped into a Stimulus space. In all the experiments, there are always both individuals and stimuli—what is it that determines whether an experimenter maps his experiment into a Stimulus space or into a Joint space?

If you go back and look at experiments with this question in mind, it becomes obvious that the experimenter puts his experment in a Joint space or a Stimulus space according to whether he regards the individual as having evaluated the stimuli with respect to a point corresponding to himself, the individual, or whether the individual evaluated the stimuli with respect to an attribute. I have called these two kinds of tasks, task A and task B, respectively. Task A may be described as evaluative, having to do with the relation of stimuli to the individual himself. Task B may be described as substantive, having to do with the nature of the stimuli per se.

I formalized this distinction between task A and task B in the following manner. In all experiments, both the individuals and the stimuli are points in a space, but in task A the points associated with the individuals are independent of the points associated with the stimuli. Whereas in task B, where the individual is evaluating stimuli with respect to an attribute, the point associated with an individual is completely dependent upon the points associated with the stimuli he is evaluating. I will not try to go further with this now, but essentially what we have is one dimension of the Scale Grid with just its two extremes represented—complete independence of the individual's point from those of the stimuli and complete dependence of his point on those of the stimuli, corresponding to task A and task B, respectively.

I have taken a good deal of time just to give an intuitive notion of one dimension of the Scale Grid. Let me briefly say just a few words about a second dimension. We can ask exactly the same questions about the difference between a Joint space and a Population space; we would find an equivalent answer if we just reverse the roles of stimuli and individuals in the argument and analysis just made. The reasoning is not difficult but is too detailed for an address. Let me merely state the conclusions. In a Joint space the points associated with the stimuli are completely independent of the points associated with the individual, whereas in a Population space the points associated with the stimuli are completely dependent on the points of the individuals responding to them. There is a duality between Stimulus spaces and Population spaces: in Stimulus spaces the points associated with individuals are dependent upon the points associated with the stimuli judged, whereas the reverse holds for Population spaces.

We have here two of the dimensions of the Scale Grid. I have constructed

two others which can be used to characterize the data within these major areas of Joint spaces, Stimulus spaces, and Population spaces; but I will say nothing further about them here as they are not relevant to the interpretations I wish to bring out. I will only say that something of their nature is described in my early monograph on theory of psychological scaling (**2**).

If we take the two dimensions which we already have, they suggest a fourth type of space, called a *Field* space, in which the points associated with stimuli and those associated with individuals are completely mutually dependent. This gives us a two-dimensional Scale Grid as illustrated in Figure 5.

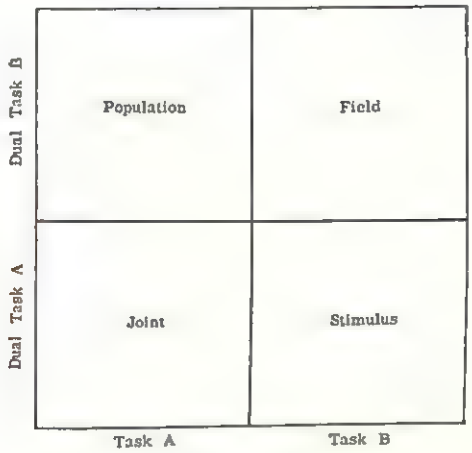


FIGURE 5

The Scale Grid

Some questions naturally arise as to just what might be the significance of all of this and just what this Field space type of behavior is. While we know what goes into a Stimulus, Population, or Joint space, this Field space was a consequence of our analysis of the others, and it is not immediately obvious what significance it has. In order to answer this question, I shall suggest a psychological interpretation of the Scale Grid.

Psychological Interpretation of the Scale Grid

I shall first point out what I consider to be the psychological processes involved in collecting data in a Stimulus space. Then by virtue of the duality between stimuli and individuals, a dual interpretation for Population spaces will be made and certain implications of duality pointed out. From these two kinds of spaces we shall move in one direction and get Joint spaces and in the other direction to get Field spaces.

Consider what is involved when we collect data in a Stimulus space. An individual, the subject, is asked to make judgments about stimuli with respect to an attribute (task *B*). He is given a set of weights and asked about their

felt-heaviness, a set of tones and asked about their pitch or loudness, etc. The objects of judgment in this situation have many measures; each object has a measure on each of its many attributes—e.g., color, size, heaviness, form, volume, aesthetic quality, etc.

So we have stimuli corresponding to points having several components, and individuals instructed to select one of these components and evaluate the stimuli with respect to that attribute. Let me say it again and contrast the difference—the stimuli are points on many attributes, the individual comes with an attribute in mind but no scale position of his own from which to evaluate the stimuli. In an exaggerated sense, for a Stimulus space the stimuli have provided the points, and the individual has provided the attribute. These then are the respective functions of stimuli and individuals in generating data in a Stimulus space. The behavior observed is ultimately converted to measures of the stimuli on the attribute. The behavior observed may run the gamut of paired comparisons, rating scales, or free-answer protocols—these differences are not relevant here. The important thing is that the behavior observed is interpreted as data which are relations on the stimuli with respect to the attribute. The analysis then leads to measures of the stimuli only.

Let us now exercise the duality relation and consider what the process must be for a Population space. When we reverse the roles of stimuli and people we must have a group of individuals, each possessing measures on many attributes just as the stimuli had in a Stimulus space. An individual thus corresponds to a point with many components. Then we must have stimuli which, carrying through the analogy, must be instructed to select one of these components and evaluate the individuals with respect to that attribute. In the Population space the individuals provide the points, and the stimulus provides the attribute. These then are the respective functions of stimuli and individuals in a Population space. What kinds of behavior do psychologists observe in which these are the respective functions of the stimuli and the subjects? I would say that certain questionnaires, certain interest and neurotic inventories, and essay examinations represent the kinds of behavior that are typically mapped into Population spaces. There is a variety of possible methods of observing such behavior, but, speaking category-wise, the most typical are rating scales and free-answer protocols.

The rating scale method is the questionnaire item with a number of ordered alternatives—an example from a questionnaire used on soldiers during the war is the following:

Do you ever get so blue and discouraged that you wonder whether anything is worth while?

a) Hardly ever
b) Not so often
c) Pretty often
d) Very often

Such a procedure is formally equivalent to asking an individual to judge weights as being light, medium, or heavy. But one experiment is in a Population space, the other in a Stimulus space.

Free-answer protocols are illustrated by the essay examination, the open-ended questionnaire, and the interview. The individual is asked a question which in principle specifies an attribute, e.g., How do you feel about the farm policy? The individuals who are asked this question are playing the role of stimuli being evaluated with respect to an attribute. The protocols which emerge are analyzed for relations between the individuals, which constitute measures of them on this attribute. This is formally equivalent to the use of individual's evaluations of stimuli with respect to an attribute, which leads to information about where the stimuli are on the attribute selected by the individual.

One immediate consequence of this duality between Stimulus spaces and Population spaces is that any method for collecting or analyzing data constructed for either one of these spaces immediately becomes a potential method for collecting or analyzing data for the other. Thus we have, for example, Thurstone's Law of Comparative Judgment constructed for analyzing the judgment of individuals about stimuli and arriving at a Stimulus scale. Immediately there is implied the dual method of having stimuli make paired comparison judgments about individuals—as yet I have seen no good way of getting stimuli to do this. However, there are variations of this basic method of Thurstone's: the Method of Successive Intervals and the Method of Equal Appearing Intervals are used for constructing Stimulus spaces which do transfer completely to Population spaces. To transfer the Method of Successive Intervals to Population spaces you need to have stimuli sort people into piles. I suggest this is exactly what is done by those questionnaire items with multiple alternatives, e.g., from strongly agree to strongly disagree. Abstractly, we can look upon such behavior as stimuli sorting individuals into piles. In the Method of Successive Intervals the instructions to the subject are dualistically equivalent to the writing and editing of an item for a questionnaire or essay examination. I find it strikingly curious that we frequently tend to use five degrees of indorsement or five ordered steps in the alternatives, whereas in the Method of Successive Intervals we have individuals sort items into as many as eleven piles. Whether there is a profound reason behind this, or whether it is unjustified adherence to tradition, I am not sure.

Just as Thurstone's methods for Stimulus scales are transferable to Population spaces, so also are methods designed for the analysis of data in Population spaces alternative methods for Stimulus spaces, e.g., Lazarsfeld's methods of latent structure analysis could be used for scaling stimuli in Stimulus spaces by reversing the subscripts which identify stimuli and individuals, and obtaining the appropriate kind of judgments from individuals.

We see here, in fact, the relation of certain methodologies of the psycho-

physicist studying Stimulus spaces to those of the social psychologist studying Population spaces. What the first makes people do to the stimuli, the latter makes stimuli do to people. Their methods of doing research, collecting, and analyzing data are formally isomorphic with the roles of stimuli and people reversed. Surely, with reference to methodology, whatever one develops suggests a dual development for the other. Each delineates an attribute on which the objects of judgment are to be evaluated. In psychophysics the experimenter does this through his instructions to the subject; in questionnaires and essay examinations the experimenter does it through his careful writing of items. All the various experimental controls developed in one context, again, in principle, transfer to the other context with the reversal of roles between stimuli and people.

The greater status in measurement of psychophysics is due, at least in part, to the fact that an individual can compare two stimuli directly, whereas a stimulus cannot compare two individuals directly. We are much happier with the judgment of an individual as to which of two attitude statements he prefers to indorse than we are with the judgment of which of two individuals indorses a given statement more strongly. The reason for this is very simple. When we ask an individual which of two stimuli he prefers, we assume he has an implicit standard of measurement that is an ordered scale applicable to both stimuli. If we wish to compare two individuals as to which indorses a stimulus more strongly, we have to assume not only that they each have an implicit interval scale but also that the scales have the same origin and the same unit of measurement. Thus, if individual A says he would pay \$10 for a picture and individual B \$5, how do we know but that A has less value for money than B? Once we have made the assumption of an interpersonally comparable interval scale, there is no sense in reducing the data to a paired comparison—that would be throwing away information already assumed. This argument can be summed up by saying that the implicit standards of judgment of one person are presumably more stable over the two stimuli than the standards of two people over one stimulus. This might well be the consideration that underlies using fewer alternatives for an item in a questionnaire than the number of piles used in the Method of Successive Intervals.

When one looks at the differences between areas in this context one finds no justification for quarrels nor for differences in respectability. One area can use a system just as logically precise as the other, but the basic data observed in one area may be a weaker relation than is observed in the other area.

With Stimulus and Population spaces mutually described and related we turn briefly to Joint spaces. Here both stimuli and individuals come together, jointly specifying what the attributes will be; both have their own measures on these attributes. For example, consider an individual working an arithmetic problem. The arithmetic problem is represented by

a point with measures on one or more components. The individual is similarly represented by a point with measures on these components. The response of the individual to the stimulus will be information about the relation of these two points in the Joint space.

By defining what this information is in different ways, one gets Guttman scalogram theory, test theory, one of Lazarsfeld's models, or my unfolding technique for the analysis of preferences. All of these are just different models for what a response on the phenotypic level means in terms of distances between pairs of points in the Joint genotypic space. [The relation of these various spaces to the classification of methodologies contained in (3) should be pointed out. The methods of collecting data which apply to Joint spaces are classified in Quadrants I and II, and the methods which apply in Stimulus spaces or in Population spaces are classified in Quadrants III and IV.]

In Joint spaces all the psychophysical methods for analyzing experiments concerned with thresholds, as distinct from those concerned with measuring only the stimulus magnitudes, are present. The individual in such experiments is regarded as having a threshold on an attribute, such as his sensitivity to light or his ability to discriminate pitch. This characteristic of the individual corresponds to a point in the genotypic space, which we have called his ideal on that attribute. The stimulus then is an increment of light or a difference between two tones, and the individual is asked whether he observes it. The stimulus is then also represented by a point in the space. The response of the individual is a formal relation on the pair of points in exactly the same manner as passing or failing an arithmetic item.

The data obtained from neurotic inventories and interest inventories are typically mapped into Joint spaces. For example, an individual is asked a question like "Are you shy?" which he is to answer *yes* or *no*. The individual is presumed to prossess and recognize his particular amount of shyness—this corresponds to a point in the genotypic space, which is his ideal. The question "Are you shy?" with the alternatives *yes* or *no* also corresponds to a point in the genotypic space which is that amount of shyness the individual feels he should have to say *yes*. This amount of shyness is formally equivalent to the difficulty of an arithmetic problem, the increment of light, or the difference between two tones in the preceding examples. Again the individual's response to the question is interpreted as a relation between the respective points.

That the data obtained from individual's indorsements or preferences between attitude statements may also be mapped into Joint spaces is too obvious to need further description. Most learning experiments are mapped into Joint spaces. A conditioning experiment, for example, is like an objective test given backwards: a combination of unconditioned stimulus and conditioning stimulus may be thought of as an item, and eliciting a conditioned response is equivalent to passing an item. Then the most difficult item in

the test is presented first, i.e., the first presentation of conditioning stimulus and unconditioned stimulus, and the individual usually fails it. As learning takes place each successive presentation is essentially an easier item until items are so easy that the individual passes them all. It is interesting to note that the conditioning test has a different method of scoring from the objective mental test, e.g., the number of items taken to reach a certain number of items passed successively. One wonders why most objective tests should have a different convention. I do not object to different conventions, I just like to know what the logic behind them is.

So we have all these superficially different kinds of behavior: objective tests, certain psychophysical experiments, neurotic inventories, interest questionnaries, attitude scale studies, and conditioning experiments. All tend to develop their own methodologies and their own vocabulary—but all are formally isomorphic and hence their methodologies transferable from one to the other. A model for analysis of one of these kinds of data with a particular distance function, for example, immediately raises the question whether it does not also constitute a theory about each of the other seemingly different kinds of behavior. There are, of course, differences in the characteristics of the data one gets in these areas. In some areas experimentally independent replication is possible, in others not; in some areas the stability of a point associated with an individual or with a stimulus is greater or less than in other areas. But fundamentally these differences are quantitative, not qualitative, and the methodological contributions to any one area are in principle transferable to all the others.

Before going on to a psychological interpretation of Field spaces, I should digress for a moment and clear up a possible source of confusion. I have covered Stimulus, Population, and Joint spaces using repeated illustrations. There is a danger of certain misconceptions arising from the illustrations, which we must try to avoid. When I have illustrated one of these spaces, I have tried to follow the most conventional ways of analyzing such behavior, but the implication should *not* be drawn that the theory says there is only one kind of quantitative data or only one space into which any particular behavior can be mapped.

The act of a psychologist in putting his experimental data into one of these spaces (Joint, Stimulus, or Population) represents an optional decision on his part. The sense in which these decisions are optional is what I now want to make clear. The same behavior may be put into more than one of the spaces, thus reflecting different points of view or problems in the mind of the experimenter. The distinction between behavior and data, which was made earlier, is the relevant principle here. It is sometimes easy to see how the same behavior may be interpreted separately as two different kinds of data and consequently be put into different spaces. We have become so accustomed to certain conventions in the converting of manifest behavior

into data that we sometimes neglect any mapping but the conventional one. While data is obtained from behavior by interpreting the behavior as a relation between points, it is up to the interpreter to decide what to identify as points and to define the properties of the relation.

Consider a study on nationality preferences. Let each subject make paired comparison judgments as to which nationality he prefers. In Thurstone's well-known study (**5**) such data were analyzed by the Law of Comparative Judgment and a scale obtained with the stimuli ranging on a one-dimensional continuum from most preferred to least preferred by the group as a whole. When the experimenter does this, he is regarding "preferability" as an attribute of stimuli and is saying that the individuals made task *B* judgments, substantive judgments, about the stimuli. The behavior is interpreted as an order relation on pairs of points, both of which are stimuli. He is saying the behavior belongs in a Stimulus space, and proceeds to construct a stimulus scale.

On the other hand, one could take the identical experiment and put it in a Joint space. In doing this, one would be assuming that the individuals were also points in the space and that their behavior is to be interpreted as an order relation on distances of the stimuli's points from the individual's point. Thus, the behavior is being put into a Joint space instead of a Stimulus space, and analysis of the data by multidimensional unfolding would yield a solution with both stimuli and individuals in the space.

There is nothing intrinsically correct about one of these procedures or wrong about the other. In the first instance, analyzing the data in a Stimulus space, one's problem is essentially that of amalgamating the preferences of individuals to arrive at a single preference scale, which in some sense best represents all the individuals. This is the problem of social utility or social choice. In the second instance, analyzing the data in a Joint space, one's problem is that of discovering the latent attributes underlying nationality preferences from which an individual's preferences could be derived. It might be parenthetically remarked that these two solutions would bear a certain interesting relation to each other, this relation has been developed in two previous publications (**2**, **4**).

Here we have taken an example of behavior and made the transition into two different kinds of data, analysis of which yields different results. We usually overlook this step that we take between behavior and data because this step, at least in some areas of research, is so conventional and immediate. Everyone can usually agree on what is or is not the right answer to an arithmetic problem; when an individual says this weight is heavier than that one, everyone usually agrees he means this weight is higher on an attribute of felt-heaviness than that one. But when we have an individual's answer to an essay examination question or his answer to an open-ended questionnaire item, we speak of "coding" them. This is the process of con-

verting the behavior to data by processing it through the mind of another person to get statements of magnitude or relations—these data are what are analyzed. It is important to note that what one analyzes is always data, not behavior.

This distinction between behavior and data now becomes an even more important and relevant distinction as we turn to a psychological interpretation of Field spaces. To arrive at this interpretation we move along two dimensions of the Scale Grid simultaneously. In going from Joint to Stimulus spaces, the point associated with the individual became dependent on the stimuli being judged. There ceased to be a unique point characterizing the individual. Another way of looking at it is that the stimulus ceased to define the attribute with respect to which the judgments were made. In passing from Joint to Population spaces, the point associated with a stimulus became dependent on the individuals being judged. There ceased to be a unique point characterizing the stimulus. Another way of looking at this is that the individual ceased to define the attribute with respect to which the judgments were made.

In Joint spaces both stimuli and individuals are points and jointly define the attribute. In Stimulus spaces the stimuli are independent points, and the individuals are instructed to define the attribute. In Population spaces the individuals are independent points and the stimuli are instructed to define the attribute. Putting these together for Field spaces, we have the points for stimuli and individuals mutually dependent with neither instructed to define an attribute.

If you wanted to observe such behavior what would you do? You would present an individual with a stimulus that was of such an ambiguous nature it would not arouse any common attribute space in individuals. At the same time the individual would be totally uninstructed to respond with respect to any particular attribute space. This is my definition of what would be a perfect projective test situation. The behavior that is observed is associated with a point in a psychological space with which both the individual and the stimulus are identified.

It is to be noted that in all the other types of spaces (Joint, Stimulus, and Population) the attribute space is at least implicitly defined by the stimuli and/or by the instructions to the subject. Consider all the care given to selecting and wording items properly so that they will ask exactly the right question. This is nothing more than trying to limit the attribute space generating behavior. Exactly the same objective underlies the care in the communication of instructions to the observer in a psychophysical experiment. This care is taken to insure that he will ask the same question of every stimulus. It is then assumed that these precautions have succeeded, and the behavior is interpreted as information about a pair of points, or the distance between them, or a pair of distances. This mapping, done by definition, is what translates behavior into data.

In a Stimulus space, the individual was instructed to ask of the stimulus how heavy it was, or how esthetically pleasing it was, etc. An attribute was explicit, and so the behavior could be interpreted as magnitudes on an attribute and thus made into data. Analysis of such data leads to conclusions about *interstimulus differences*. In a Population space the roles are reversed: a stimulus comes to the individual and asks him how he feels about the farm policy. Again an attribute is explicit. The behavior is interpreted as magnitudes on an attribute, analysis of which leads to conclusions about *interindividual differences*.

In a Joint space both interpretations are possible because the behavior is interpreted as data on a relation between individuals and stimuli. In a Field space the point associated with the individual has merged with the point associated with the stimulus. The behavior is information about this point in a psychological space, a point in which the subject and the stimulus are inextricably identified.

The care given in Population spaces to selecting items for a questionnaire or essay examination in order to ask every individual the same question, and the care given in Stimulus spaces to phrasing instructions to the subject in order that he evaluate all the stimuli on the same attribute is now exercised in Field spaces so that precisely these effects will *not* occur. Every effort is made to insure a setting in which the stimulus will *not* suggest a particular attribute space, and every effort is made in the instructions to the subject *not* to suggest a particular attribute space. Herein lies both the strength and weakness of Field spaces. The protocol that emerges now constitutes a stimulus to be evaluated—so it is a stimulus to be located in a Stimulus space. In order to convert this protocol, this behavior, into data certain problems need to be solved. One is: what is the attribute or attributes which underlie the behavior? This is a new problem which had not previously arisen for any of the other spaces. Here now we have a protocol which is to be converted into a measure on some attributes. The first problem is: which attributes? This problem arises because there was not deliberately built into the stimulus nor into the individual constraints or instructions which would provide a simple answer.

It is immediately obvious that behavior in this area does not lead to interindividual comparisons, because there has been no instruction to the stimulus, no built-in device by virtue of which it can be assumed that a stimulus has evaluated each individual on the same attribute. If one person exhibits guilt feelings and another does not, one cannot conclude the latter person has less guilt feelings unless one can assume the stimulus was such as to make every individual reveal his guilt feelings. In this case, of course, such data could be put back in Population spaces, and interindividual comparisons would be possible.

It may help one to recognize and understand this problem if we point

out that it is like having the answer of an individual to an essay examination question when you do not know what the question was. Consider, then, having the answers of several individuals, each to an unknown question; the problem is to decide which individual's answer represents more of someing than another's. It seems to me the problem is meaningless if the question answered by an individual has been left up to him to select, hence each individual has perhaps selected a different question. One could say, well I can evaluate their relative command of English, or their vocabulary level, or their handwriting if it is a written protocol. This is entirely correct, of course, and amounts to saying this is the common attribute which the stimulus aroused in all individuals, hence they may legitimately be compared. This puts the behavior into a Population space, not a Field space. What I am talking about are those aspects of the behavior which are not attribute controlled and hence belong in a Field space.

Understandably enough, there have been instruments constructed, called projective instruments, which seek to avoid this particular problem. For instance, there are test instruments in which a picture suggests an attribute; the individual is asked to write a story which is presumed to reveal where he is on that attribute. Examples are Proshansky's Labor TAT, Johnson's Anglo-Spanish TAT, and toy play with negro and white dolls. If these instruments succeed in their purpose, then we have the stimulus coming to the individual with an attribute in mind and asking the individual where he is on it. These instruments then, if successful, are formally the same thing as a rather subtle essay examination or an interview by a laborer, a Mexican, or a Negro. The data that are obtained pertain to a Population space rather than a Field space, and interindividual comparisons are logically permissible. Such instruments, however, are not projective instruments in the sense of belonging in Field spaces. A further question then arises as to whether or not these instruments accomplish their purpose. If they fail to arouse the attribute which the experimenter subtly built into the stimulus then there is serious danger of drawing false conclusions.

Let us assume that the first problem is solved or can be solved—that we can look at the protocol and say what the attributes are which underlie the behavior. Then a second problem arises, which is the most fascinating and perhaps the most important of all: what does it mean that the individual selected these particular attributes to exhibit out of all of those possible in his repertoire? I think that no solution yet exists, but ultimately this problem must be answered in order to interpret projective instruments. This problem lies in the area of the psychology of the individual. Because the attributes were left up to the individual, their selection is a reflection of his internal dynamics. Because each individual is answering different questions, the behavior cannot be taken to reflect interindividual comparisons on a common

attribute. On the contrary, and therein lies both its importance to psychology and its weakness as data, the behavior reflects *intra-individual* comparisons.

A protocol in a Field space reflects a point in a psychological space. When we know enough to interpret the protocol as a measure of that point in a known attribute space, then we shall be able to make comparisons between the points. I suspect that these will be comparisons on some hyperabstract attributes which will reflect intra-individual dynamics. The problems which must be solved to reach this stage are what I would regard as our ultimate measurement problems. Field spaces are a maximally significant domain of behavior. It is the area that reflects intra-individual differences to a degree that no other area does. There are fascinating and important problems for psychologists here. It is my thought and hope that the Scale Grid will help to delineate more clearly the basic measurement problems involved.

REFERENCES

1. Bentley, M. Early and late metric uses of the term "distance." *Am. J. Psychol.*, 1950, **63**, 619.
2. Coombs, C. H. A theory of psychological scaling. *Eng. Res. Inst. Bull. No.* 34, Ann Arbor, Mich.: Univ. Mich. Press, 1952.
3. Coombs, C. H. Theory and methods of social measurement. In L. Festinger and D. Katz (eds.), Research methods in the behaviorial sciences. New York: Dryden Press, 1953.
4. Coombs, C. H. Social choice and strength of preference. In R. M. Thrall, C. H. Coombs, and R. L. Davis (eds.), Decision processes. New York: Wiley, 1954.
5. Thurstone, L. L. An experimental study of nationality preferences. *J. Gen. Psychol.*, 1928, **1**, 405-424.

Manuscript received 9/4/56

PSYCHOMETRIKA—VOL. 21, NO. 4
DECEMBER, 1956

EFFICIENT ESTIMATION AND LOCAL IDENTIFICATION IN LATENT CLASS ANALYSIS*

RICHARD B. MCHUGH

IOWA STATE COLLEGE

Estimators which are efficient in the sense of having minimum asymptotic variance are obtained for the structural parameters of Lazarsfeld's latent class model of latent structure analysis. Sufficient conditions for the local identification of the structural parameters are also presented.

1. *Introduction*

A psychological test of p items or subtests will be denoted by $(x_1, x_2, \cdots, x_p)$. Data obtained from the administration of a psychological test can be processed in different ways. For example, in Lazarsfeld's latent structure analysis (14) the test data are treated qualitatively, i.e., as measurements on a nominal or at most an ordinal scale. On the other hand, in classical Spearman-Thurstone factor analysis test data are treated quantitatively, i.e., as measurements on an interval scale. Despite this difference there is a feature common to these two approaches, viz., the view that the αth examinee's observed test behavior $(x_{1\alpha}, x_{2\alpha}, \cdots, x_{p\alpha})$ may be thought of as being determined by his status $(\xi_{1\alpha}, \xi_{2\alpha}, \cdots, \xi_{\lambda\alpha})$ on λ traits or factors $(\xi_1, \xi_2, \cdots, \xi_\lambda)$ underlying the test. That is, observed test behavior (referred to as phenotypic or manifest) is viewed as only an indicant of, not as a direct measurement of, the underlying traits (referred to as genotypic or latent). [Among others, Coombs (2) and Stevens (13) offer justification for this approach to psychological measurement.]

Fundamental to this kind of test analysis are the concepts of *structure*, *model*, and *identification*. In section 3 below, the general formulation of these concepts due to Hurwicz (5), to Koopmans and Reiersøl (6), and to Reiersøl (11) is applied to a particular model of Lazarsfeld's latent structure analysis, viz., the latent class model.

In addition to (*a*) the pre-statistical problem of identification of the model, other problems which arise in this approach to the analysis of a psychological test include (*b*) the efficient estimation of the identifiable parameters of the model, and (*c*) the scaling problem, i.e., the assignment of trait measurements $(\xi_{1\alpha}, \xi_{2\alpha}, \cdots, \xi_{\lambda\alpha})$ to each examinee. Section 3 is con-

*The writer wishes to acknowledge with appreciation the helpful advice of Professors Leonid Hurwicz and Jacob Bearman of the University of Minnesota. Dr. John Gurland of Iowa State College assisted the author in clarifying certain points in this article.

cerned with problem (*a*) and section 4 with problem (*b*). [Problem (*c*) has been discussed for the latent class model by Lazarsfeld (14)]. Section 5 presents a statistical test of the goodness of fit of the structure, and in section 6 a numerical example is presented to illustrate the mechanics of computation.

Section 2 provides a resumé which makes certain features of Lazarsfeld's presentation (14) explicit in a form and notation convenient to the remainder of the discussion.

2. *The Latent Class Model*

Three distribution functions and an assumption concerning one of them constitute the essential ingredients of the latent class model.

(*i*) Let $G(x_1, x_2, \cdots, x_p)$ denote the distribution function of the p observable (manifest) random variables corresponding to the p items or subtests of the test. In practice, the subtests or items are usually dichotomous, i.e., the response of the αth subject to x_i, denoted by $x_{i\alpha}$, equals 1 for a *positive* response and 0 for a *negative* response ($i = 1, 2, \cdots, p$). Since the domain of the ith random variable, x_i, is 0 and 1, then $G(x_1, x_2, \cdots, x_p)$ is a p-dimensional binomial distribution function. Let the set of parameters defining the test response distribution, $G(x_1, x_2, \cdots, x_p)$, be denoted by $\{g_{i_1 i_2 \cdots i_p}\}$. These parameters are called *manifest marginal* probabilities. There are 2^p of the elements $g_{i_1 i_2 \cdots i_p}$, viz., $g_{12\cdots p}, g_{\bar{1}2\cdots p}, \cdots, g_{\bar{1}\bar{2}\cdots\bar{p}}$. By definition, $g_{12\cdots p} = P\{x_1 = 1, x_2 = 1, \cdots, x_p = 1\}$, $g_{\bar{1}2\cdots p} = P\{x_1 = 0, x_2 = 1, \cdots, x_p = 1\}$, and so forth to $g_{\bar{1}\bar{2}\cdots\bar{p}} = P\{x_1 = 0, x_2 = 0, \cdots, x_p = 0\}$.

In the latent model, then, the test may be interpreted as a random vector $(x_1, x_2, \cdots, x_i, \cdots, x_p)$ of p components—no effort is made to quantify the item response patterns. This may be contrasted with the mental test theory models, Gulliksen (4), Lord (7), where the test results are summarized in a total score, $\sum_{i=1}^{p} x_i$. The mental test theory models, therefore, use less of the information available from the test behavior than does the latent class model. In a given test situation, the loss in information may not be of practical significance if p is large, whereas the gain in computational convenience may be practically significant.

(*ii*) Let $F(\xi_1, \xi_2, \cdots, \xi_\lambda)$ denote the distribution function of the λ unobservable (latent) random variables. However, since the aim in practice (14) is generally to construct a *pure* test, i.e., a test measuring a single trait, it will be assumed here that $\lambda = 1$. That is, although the test in (*i*) is conceived as consisting of p variables, the trait ξ is postulated to be a one-dimensional random variable with distribution $F(\xi)$.

The domain of ξ need not be restricted to a point set of two elements, 0 and 1, as is the domain of x_i. In general, the domain of ξ consists of a point set of γ elements. These elements are termed *latent classes*. The generic class will be denoted by c ($c = 1, 2, \cdots, \gamma$). For example, in the numerical illustration in section 6, where the latent trait ξ is creativity in machine

design, the case of $\gamma = 2$ is considered, viz., latent class $c = 1$ of creative machine designers and latent class $c = 2$ of non-creative machine designers. (By means of a chi square test of goodness of fit, however, as noted in section 6, a possible inference for this example is that $\gamma > 2$).

On the basis of this qualitative interpretation of the latent trait, $F(\xi)$ is a one-dimensional multinomial distribution function, specifically a one-dimensional γ-nomial. Let the set of parameters defining the latent trait distribution $F(\xi)$ be denoted by $\{f_c\}$. These parameters are termed *latent marginal* probabilities. There are γ of the elements f_c, viz., $f_1, f_2, \cdots, f_\gamma$. By definition, $f_c = P\{\xi = c\}$, $(c = 1, 2, \cdots, \gamma)$.

For this model, the conceptual measurement ξ_α of ξ for the αth examinee (i.e., ξ_α is conceived of as the value that would be obtained for the αth subject if the latent trait ξ could be measured directly) signifies an assignment of the αth subject to one of the latent classes $c = 1, 2, \cdots, \gamma$. This problem of scaling, i.e., of drawing a latent inference from the observed test response to the unobserved trait, is problem (*c*) cited in section 1.

(*iii*) The final distribution function of the latent class model is the distribution of the subtests $x_1, x_2, \cdots, x_p$ for fixed ξ. Let $G(x_1, x_2, \cdots, x_p \mid \xi = c)$ denote this conditional distribution function.

This conditional distribution is important because it makes explicit the relation between the observable variables x_i taken jointly and the unobservable ξ. For example, one of the 2^p parameters defining this p-dimensional binomial distribution $G(x_1, x_2, \cdots, x_p \mid \xi = c)$ is the probability of a conditional joint positive response, say $g_{12\cdots p|c}$. Defined for the αth subject, this is $g_{12\cdots p|c} = P\{x_{1\alpha} = 1, x_{2\alpha} = 1, \cdots, x_{p\alpha} = 1 \mid \xi = c\}$, indicating that the αth subject's test behavior depends probabilistically upon his status on the trait ξ.

Instead of dealing with the joint conditional distribution $G(x_1, x_2, \cdots, x_p \mid \xi = c)$, however, it may suffice to deal instead with the marginal conditional distributions $G(x_1 \mid \xi = c), G(x_2 \mid \xi = c), \cdots, G(x_p \mid \xi = c)$. This great simplification is possible if the psychological test has been constructed in such a way that

$$G(x_1, x_2, \cdots, x_p \mid \xi = c) = G(x_1 \mid \xi = c)G(x_2 \mid \xi = c) \cdots G(x_p \mid \xi = c), \tag{2.1}$$

i.e., in case the subtests or items are jointly statistically independent for fixed trait level or latent class $\xi = c$. For a specified population of examinees, a test with this desirable property is called *pure* (**14**). That is, if (2.1) holds, each x_i involves a single common trait ξ. Without this factorability condition, the subtests or items would involve, ambiguously, several common traits $\xi_1, \xi_2, \cdots, \xi_\lambda$ $(1 < \lambda \leq p)$. In all that follows, Lazarsfeld's postulate (2.1) of a pure test will be assumed. The left and right sides of (2.1) will therefore be used interchangeably.

Because of their importance in the latent class model, the conditional distribution functions $G(x_i \mid \xi = c)$ of the various items, or subtests, are given the special title of *item trace functions* (**14**). These $G(x_i \mid \xi = c)$ are one-dimensional binomial distributions. By definition, the generic defining parameter is $g_{i;c} = P\{x_i = 1 \mid \xi = c\}$. The $g_{i;c}$ are referred to as *item conditional probabilities.*

To summarize, the latent class model involves the three distributions:

(*i*) $G(x_1, x_2, \cdots, x_p)$, defined by the set of 2^p parameters $\{g_{i_1 i_2 \cdots i_p}\}$;
(*ii*) $F(\xi)$, defined by the set of γ parameters $\{f_c\}$;
(*iii*) $G(x_1, x_2, \cdots, x_p \mid \xi = c)$, defined by the set of $p\gamma$ parameters

$$\{g_{i;c}\}, \qquad (i = 1, \cdots, p; c = 1, \cdots, \gamma).$$

TABLE 1

Schematic Representation of the Latent Class Model

One latent variable		ξ			
Domain of ξ (i.e., γ latent classes)		1	2 ...	c ...	γ
Latent marginal probabilities		f_1	$f_2 \cdots$	$f_c \cdots$	f_γ
p manifest variables (items or subtests) x_i	Domain of x_i (i.e., dichotomy)	Item conditional probabilities			
x_1	1	$g_{1;1}$	$g_{1;2} \cdots$	$g_{1;c} \cdots$	$g_{1;\gamma}$
	0	$g_{\bar{1};1}$	$g_{\bar{1};2} \cdots$	$g_{\bar{1};c} \cdots$	$g_{\bar{1};\gamma}$
x_2	1	$g_{2;1}$	$g_{2,2} \cdots$	$g_{2;c} \cdots$	$g_{2;\gamma}$
	0	$g_{\bar{2};1}$	$g_{\bar{2};2} \cdots$	$g_{\bar{2};c} \cdots$	$g_{\bar{2};\gamma}$
⋮	⋮	⋮			⋮
x_i	1	$g_{i;1}$	$g_{i;2} \cdots$	$g_{i;c} \cdots$	$g_{i;\gamma}$
	0	$g_{\bar{i};1}$	$g_{\bar{i},2} \cdots$	$g_{\bar{i};c} \cdots$	$g_{\bar{i};\gamma}$
⋮	⋮	⋮			⋮
x_p	1	$g_{p;1}$	$g_{p;2} \cdots$	$g_{p;c} \cdots$	$g_{p;\gamma}$
	0	$g_{\bar{p};1}$	$g_{\bar{p};2} \cdots$	$g_{\bar{p};c} \cdots$	$g_{\bar{p};\gamma}$

Table 1 gives a convenient scheme for representing $F(\xi)$ and $G(x_1, x_2, \cdots, x_p \mid \xi = c)$, as defined by $\{f_c\}$ and $\{g_{i;c}\}$, respectively.

The three distributions (*i*), (*ii*), and (*iii*) are related by a set of 2^p equations for the latent class model:

$$\begin{aligned} g_{12\cdots p} &= \sum f_c g_{1;c} g_{2;c} \cdots g_{p;c}, \\ g_{\bar{1}2\cdots p} &= \sum f_c g_{\bar{1};c} g_{2;c} \cdots g_{p;c}, \\ &\cdots \\ g_{\bar{1}\bar{2}\cdots\bar{p}} &= \sum f_c g_{\bar{1};c} g_{\bar{2};c} \cdots g_{\bar{p};c}, \end{aligned} \tag{2.2}$$

where the summation is from $c = 1$ to γ. The set (2.2), the accounting equations, follows from the fact that the distribution $G(x_1, x_2, \cdots, x_p)$ can be uniquely determined from the distributions $F(\xi)$ and $G(x_1, x_2, \cdots, x_p \mid \xi = c)$ by summing the product $F(\xi)G(x_1, x_2, \cdots, x_p \mid \xi = c)$ over the domain of ξ.

It should be noted that of the 2^p manifest marginal probabilities only $2^p - 1$ can be functionally independent since

$$\sum_D g_{i_1 i_2 \cdots i_p} = 1, \tag{2.3}$$

where the summation is over all the 2^p values of $i_1 i_2 \cdots i_p$, a domain which will be denoted by D. Similarly, of the latent marginal probabilities only $\gamma - 1$ can be independent because

$$\sum_{c=1}^{\gamma} f_c = 1.$$

Finally, the item conditional probabilities are related in general as

$$g_{i;c} + g_{\bar{i};c} = 1.$$

3. *Local Identification of the Structural Parameters f_c and $g_{i;c}$*

The parameters f_c and $g_{i;c}$ ($i = 1, \cdots, p$; $c = 1, \cdots, \gamma$) of Table 1 are called *structural parameters* for the following reason. A structure, denoted here by $S^0 = \{F^0(\xi), G^0(x_1, x_2, \cdots, x_p \mid \xi = c)\}$, consists of the combina tion of particular, concretely specified distributions $F^0(\xi)$ and $G^0(x_1, x_2, \cdots, x_p \mid \xi = c)$. Equivalently, in terms of the defining parameters, $S^0 = \{f_c^0, g_{i;c}^0\}$, ($i = 1, \cdots, p$; $c = 1, \cdots, \gamma$). That is, a structure is the combination of a set of particular numbers f_c^0 and $g_{i;c}^0$ for the parameters f_c and $g_{i;c}$. Table 4 presents an example of an estimated structure.

From (2.2) it is clear that a particular structure $\{f_c^0, g_{i;c}^0\}$ uniquely determines the set of parameters $\{g_{i_1 i_2 \cdots i_p}^0\}$ defining the distrubution function $G^0(x_1, x_2, \cdots, x_p)$ of the manifest subtests. The set $\{g_{i_1 i_2 \cdots i_p}^0\}$ is said to be generated by the structure S^0. Because each different (permissible) set of specific numbers f_c^0 and $g_{i;c}^0$ constitutes a different structure S^0, it is natural to consider the class $C = \{S\}$ of all such structures—this is the technical definition of a model. By definition, the *latent class model* C is the class of all structures $\{S\} = \{f_c, g_{i;c}\}$ conformable with the specifications of section 2. By contrast, an individual structure S^0 is a particular realization of the model, C.

It will be assumed in what follows that there is at least one structure. Then a problem of identification arises in a natural fashion. That is, a particular structure $\{f_c^0, g_{i;c}^0\}$ generates one and only one set of manifest parameters $\{g_{i_1 i_2 \cdots i_p}^0\}$. Hence the very practical question, called the *identification* question, is then posed, viz.: Does the converse proposition hold, i.e., can $\{g_{i_1 i_2 \cdots i_p}^0\}$ be generated by only the structure $\{f_c^0, g_{i;c}^0\}$? If so, then a given set of

Ednl. Research
COLLEGE
No

manifest marginal probabilities $\{g^0_{i_1 i_2 \cdots i_p}\}$ determines in principle one and only one structure $\{f^0_c, g^0_{i:c}\}$. If this were to occur, the latent class model $C = \{S\}$ would be said to identify (uniquely) the particular structure S^0, or the structure S^0 would be termed (uniquely) identifiable by the latent class model.

One necessary condition for identifiability of the structure S^0 or set of structural parameters f^0_c and $g^0_{i:c}$ is clearly that the number of independent pieces of manifest information be at least as great as the number of independent latent unknowns, i.e., the number, $2^p - 1$, of independent manifest parameters $g^0_{i_1 i_2 \cdots i_p}$ be at least as great as the number $\gamma - 1 + p\gamma$ of independent structural parameters f^0_c and $g^0_{i:c}$, or

$$2^p \geq \gamma(p + 1). \tag{3.1}$$

The answer to the general identification question may be that several structures besides S^0 generate the same manifest test distribution, i.e., the structure S^0 is not (uniquely) identifiable by the latent class model. However, a weaker form of identification, called *local identification*, may exist, viz., it may be that other structures generate $G^0(x_1, x_2, \cdots, x_p)$, but none of these lies in the neighborhood of S^0. (For convenience, the superscript "0" will be omitted in the following.)

THEOREM 1. *If specifications (i) through (v) hold for a structure of the latent class model, then the structural parameters f_c and $g_{i:c}$ are locally identifiable.*

(*i*) $2^p > \gamma - 1 + p\gamma$.

(*ii*) $\sum f_c g_{1:c} g_{2:c} \cdots g_{p:c} + \cdots + \sum f_c g_{\bar{1}:c} g_{\bar{2}:c} \cdots g_{\bar{p}:c} = 1$.

(*iii*) $\sum f_c g_{1:c} g_{2:c} \cdots g_{p:c} > 0$,

$$\cdot \quad \cdot \quad \cdot \quad ,$$

$\sum f_c g_{\bar{1}:c} g_{\bar{2}:c} \cdots g_{\bar{p}:c} > 0$.

(*iv*) $\sum f_c g_{1:c} g_{2:c} \cdots g_{p:c}$,

$$\cdot \quad \cdot \quad \cdot \quad ,$$

$\sum f_c g_{\bar{1}:c} g_{\bar{2}:c} \cdots g_{\bar{p}:c}$,

are continuous functions of $f_c, g_{1:c}, \cdots, g_{p:c}$ and possess continuous first and second partial derivatives.

(*v*) At least $\gamma - 1 + p\gamma$ of the expressions

$$\sum f_c g_{1:c} g_{2:c} \cdots g_{p:c},$$

$$\cdot \quad \cdot \quad \cdot \quad ,$$

$$\sum f_c g_{\bar{1}:c} g_{\bar{2}:c} \cdots g_{\bar{p}:c},$$

are functionally independent.

This theorem represents an application of a more general proposition, the heuristic origin for which is to be found in a classic paper by Fisher (3) on the limiting form of the Karl Pearson chi square criterion. A rigorous statement and proof of this more general proposition has been provided by Neyman (8, p. 250) in connection with the property of consistency of point estimators, i.e., the property of convergence in probability of an estimator to its corresponding parameter. That the proposition arises in connection with this property of estimators is not surprising since in ordinary practice only sample, not population, data are available. The following paragraph is concerned with the feasibility of the hypotheses (*i*) through (*v*) of Theorem 1 for the latent class model. This theorem is then related to the point estimation property of consistency.

Hypothesis (*i*) is equivalent to (3.1) and can be readily checked in any particular application. For example, it is satisfied in the example in section 6, where $p = 4$ and $\gamma = 2$, so that

$$2^p = 16 > 2 - 1 + (4)(2) = 9.$$

Hypothesis (*ii*) follows directly from (2.2) and (2.3). Specification (*iii*), on the other hand, cannot be verified directly; however, substantive considerations may strongly suggest it. For example, for $p = 4$ and $\gamma = 2$, if the contrary were true and

$$\sum_{c=1}^{2} f_c g_{1;c} g_{2;c} g_{3;c} g_{4;c} = 0,$$

then either $f_1 = 1$ (so $f_2 = 0$) and an even number of the $g_{i;2} = 0$, or an odd number of the $g_{i;2} = 0$—implications that are highly unlikely from the psychologist's viewpoint. (Thus, $f_1 = 1$ implies a degenerate application of the latent class model, viz., that almost all subjects belong in a single latent class—a vacuous scaling of ξ. Again, $g_{i;2} = 0$ implies that the ith subtest or item has almost no discriminating power and so is unlikely to be incorporated into the test initially.) The continuity hypothesis, specification (*iv*), clearly holds since the functions are polynomials in f_c and $g_{i;c}$. Finally, specification (*v*) is equivalent to asserting that the rank of the Jacobian matrix of the expressions is $\gamma - 1 + p\gamma$. It follows readily that full rank for the Jacobian is in turn equivalent to the requirement that the information matrix I (defined in section 6) be non-singular. This condition on I may be examined by direct algebra for a given model, or indirectly by numerical calculation of the approximate inverse of I, as in the example in section 6.

The connection between local identification of a structural parameter and the property of consistency is as follows: If it is possible to show that an estimator of the structural parameter θ, est. θ, is consistent, then θ must be locally identifiable. For example, in the latent class model if it is possible to show that the structural parameter f_c has an estimator est. f_c which

converges stochastically to f_c, then f_c must be locally identifiable. This connection may be established by noting that if a structural parameter is not locally identifiable, i.e., in a neighborhood of the parameter, the parameter is not uniquely determined from the probability distribution of the observed variables, then every estimator of this parameter will fail to converge in probability to the parameter.

In terms of estimators, the application of the Neyman general proposition (**8**) to the latent class model means that Theorem 1 may be stated as follows: If the specifications (*i*) through (*v*) hold for a structure of the latent class model, then the estimators $\hat{f}_c$ and $\hat{g}_{i;c}$ of (4.2) are consistent, i.e., the set of solutions $\hat{f}_c$ and $\hat{g}_{i;c}$ of (4.2) are such that $\hat{f}_c$ converges in probability to f_c and $\hat{g}_{i;c}$ to $g_{i;c}$ as n tends to infinity.

4. *Efficient Estimation of the Structural Parameters f_c and $g_{i;c}$*

From the p-dimensional binomial distribution of test responses $G(x_1, x_2, \cdots, x_p)$ and from (2.2) it follows that the likelihood function of the structural parameters f_c and $g_{i;c}$, based on a sample of n examinees, is

$$L(f_c, g_{i;c}) = (\sum f_c g_{1;c} g_{2;c} \cdots g_{p;c})^{n_{12\cdots p}} (\sum f_c g_{\bar{1};c} g_{2;c} \cdots g_{p;c})^{n_{\bar{1}2\cdots p}} \cdots (\sum f_c g_{\bar{1};c} g_{\bar{2};c} \cdots g_{\bar{p};c})^{n_{\bar{1}\bar{2}\cdots\bar{p}}}, \tag{4.1}$$

where $n_{i_1 i_2 \cdots i_p}$ is the number of subjects in the sample who give the response pattern $i_1 i_2 \cdots i_p$.

Application to this likelihood function of the method of maximum likelihood yields $\hat{f}_c$ and $\hat{g}_{i;c}$ as estimators of f_c and $g_{i;c}$. These estimators are the solutions of the set of equations obtained by differentiating the logarithm of (4.1), viz.:

$$\begin{aligned} &\text{(a)} \quad \frac{\partial(\log L)}{\partial f_c} = 0 \qquad (c = 1, \cdots, \gamma - 1), \\ &\text{(b)} \quad \frac{\partial(\log L)}{\partial g_{i;c}} = 0 \qquad (i = 1, \cdots, p; c = 1, \cdots, \gamma). \end{aligned} \tag{4.2}$$

Not all applications of the method of maximum likelihood result in estimators which possess the property of consistency [Neyman and Scott (**9**) give examples]. However, as noted in section 3, $\hat{f}_c$ and $\hat{g}_{i;c}$ are consistent if the specifications (*i*) through (*v*) of Theorem 1 obtain. Similarly, the routine of maximum likelihood by itself does not guarantee the property of efficiency for the resulting estimators [cf. (**9**), for examples]. In the latent class model, however, $\hat{f}_c$ and $\hat{g}_{i;c}$ are efficient if (*i*) through (*v*) obtain.

THEOREM 2. *If specifications (i)–(v) of Theorem 1 hold for a structure of the latent class model, then the estimators $\hat{f}_c$ and $\hat{g}_{i;c}$ are asymtotically efficient. That is, $\hat{f}_c$ and $\hat{g}_{i;c}$ have a joint asymptotic Gaussian distribution, and any other*

estimators of f_c and $g_{i:c}$ which are consistent and asymptotically Gaussian have asymptotic variances exceeding the asymptotic variances of $\hat{f}_c$ and $\hat{g}_{i:c}$.

[Proof of the general proposition, of which Theorem 2 is a special case, can be found in Neyman (**8**, p. 250).]

As is the case in most applications of maximum likelihood, the solutions $\hat{f}_c$ and $\hat{g}_{i:c}$ to the likelihood equations (4.2) are not in general expressible in closed form. The indirect approach of iteration must therefore be employed. That is, a trial solution $\bar{f}_c$ and $\bar{g}_{i:c}$ must be assumed and a linear system solved for small, additive corrections $\delta\hat{f}_c$ and $\delta\hat{g}_{i:c}$, after which (6.1) is applied. A convenient mechanization of this procedure, due to R. A. Fisher, is called the *scoring system*. The technical details of the scoring system are given by Rao (**10**). A numerical example is given in section 6 to illustrate the computational procedure.

5. *A Statistical Test for the Number of Significant Latent Classes.*

The final problem of statistical inference considered here is that of formulating a statistical test of the hypothesis that $\gamma = \gamma_0$, where γ_0 denotes a specified number of latent classes.

A natural approach to this problem is direct examination of the goodness of fit of the manifest marginal probabilities $g_{i_1 i_2 \cdots i_p}$, generated by the structure via (2.2), to the actual population manifest marginal parameters $g^*_{i_1 i_2 \cdots i_p}$. Thus, if the discrepancy between the generated $G(x_1, x_2, \cdots, x_p)$ and the actual $G^*(x_1, x_2, \cdots, x_p)$ is substantial, the psychologist might postulate $\gamma = \gamma_0 + 1$ in an effort to improve the fit. In practice, however, direct examination is generally impossible; statistical inference is required since only the sample estimates $n_{i_1 i_2 \cdots i_p}/n$ not the actual population parameters $g^*_{i_1 i_2 \cdots i_p}$ are available. Here $n_{i_1 i_2 \cdots i_p}$ is the sample observed response frequency corresponding to subtest response pattern $i_1 i_2 \cdots i_p$ (cf. Table 2).

For this formulation, the classical chi square goodness-of-fit test would be applicable if the structure were completely specified independently of the set of sample data $\{n_{i_1 i_2 \cdots i_p}\}$. Thus, for a postulated γ_0 and known f_c and $g_{i:c}$, the parameters $g^*_{i_1 i_2 \cdots i_p}$ of $G^*(x_1, x_2, \cdots, x_p)$ would be determined by (2.2). Hence the discrepancy between observed and theoretical frequencies, i.e., between the estimates $n_{i_1 i_2 \cdots i_p}$ of $ng^*_{i_1 i_2 \cdots i_p}$ and the $ng^*_{i_1 i_2 \cdots i_p}$ as calculated from (2.2), could be tested by:

$$\chi^2 = \sum_D \{(n_{i_1 i_2 \cdots i_p} - ng^*_{i_1 i_2 \cdots i_p})^2 / ng^*_{i_1 i_2 \cdots i_p}\},$$

which is distributed approximately as chi square with degrees of freedom $2^p - 1$.

However, in practice the structural parameters f_c and $g_{i:c}$—and therefore $g^*_{i_1 i_2 \cdots i_p}$ as obtained from (2.2)—must be estimated from the sample, hence the classical chi square test above is not valid without modification.

THEOREM 3. *If specifications (i)–(v) of Theorem 1 hold for a structure of the latent class model, then the following quantity in the limit as n tends to infinity is distributed as chi square:*

$$\chi^2 = \sum_D \{(n_{i_1 i_2 \cdots i_p} - \hat{n}_{i_1 i_2 \cdots i_p})^2 / \hat{n}_{i_1 i_2 \cdots i_p}\}, \tag{5.1}$$

with degrees of freedom $2^p - \gamma(p + 1)$, *i.e., degrees of freedom equal to the number of independent manifest parameters* $g_{i_1 i_2 \cdots i_p}$ *minus the number of structural parameters* f_c *and* $g_{i;c}$. Here $\hat{n}_{i_1 i_2 \cdots i_p}$ is the sample latent response frequency corresponding to subtest response pattern $i_1 i_2 \cdots i_p$; i.e., $\hat{n}_{i_1 i_2 \cdots i_p}$ is generated by the structure upon replacement of the structural parameters by the efficient estimates, e.g., for $p = 4$ and $\gamma = 2$ as in section 6,

$$\hat{n}_{1234} = n[\hat{f}_1 \hat{g}_{1;1} \hat{g}_{2;1} \hat{g}_{3;1} \hat{g}_{4;1} + \hat{f}_2 \hat{g}_{1;2} \hat{g}_{2;2} \hat{g}_{3;2} \hat{g}_{4;2}].$$

A rigorous generalization of this proposition is given by Neyman (8).

6. *Illustrative Example*

In order to illustrate the calculations needed, a latent class structure is estimated from data obtained by Schumacher, Maxson, and Martinek (12). Four machine design subtests, given to 137 engineers, were dichotomized into positive, 1, (above the subtest mean) and negative, 0, (below

TABLE 2

Frequency of Occurrence of Response Patterns for the Four Machine Design Subtests (12)

Response patterns $i_1 i_2 i_3 i_4$	Observed Frequencies $n_{i_1 i_2 i_3 i_4}$	Generated Frequencies $\hat{n}_{i_1 i_2 i_3 i_4}$
1 2 3 4	23	18.276211
$\bar{1}$ 2 3 4	8	8.449612
1 $\bar{2}$ 3 4	6	6.882195
1 2 $\bar{3}$ 4	5	8.915412
1 2 3 $\bar{4}$	5	8.789920
$\bar{1}$ $\bar{2}$ 3 4	9	3.676943
$\bar{1}$ 2 $\bar{3}$ 4	3	5.386429
$\bar{1}$ 2 3 $\bar{4}$	2	4.545934
1 $\bar{2}$ $\bar{3}$ 4	2	4.123700
1 $\bar{2}$ 3 $\bar{4}$	3	3.602141
1 2 $\bar{3}$ $\bar{4}$	14	5.034202
$\bar{1}$ $\bar{2}$ $\bar{3}$ 4	8	8.185339
$\bar{1}$ $\bar{2}$ 3 $\bar{4}$	3	4.205489
$\bar{1}$ 2 $\bar{3}$ $\bar{4}$	8	8.586886
1 $\bar{2}$ $\bar{3}$ $\bar{4}$	4	5.689062
$\bar{1}$ $\bar{2}$ $\bar{3}$ $\bar{4}$	34	32.650525

the subtest mean). Table 2, second column, gives the sample frequencies $n_{i_1 i_2 i_3 i_4}$ with which the observed response patterns $i_1 i_2 i_3 i_4$ occurred.

To apply the latent class model, some hypothesis concerning γ, the

number of latent classes must be made. Subsequently, this hypothesis is tested to determine whether the number is in fact sufficient to account for the data. In this example, the hypothesis is: $\gamma = 2$, i.e., that two latent classes (*Creative* versus *Non-creative* on machine design) are sufficient to account for the observed response pattern frequencies.

To start the iterative scoring system of Fisher (10) it is necessary to find first approximations $\bar{f}_c$ and $\bar{g}_{i;c}$ to the structural parameters f_c and $g_{i;c}$. Then efficient estimators $\hat{f}_c$ and $\hat{g}_{i;c}$ are given to a first iteration by

$$\hat{f}_c = \bar{f}_c + \delta\hat{f}_c \,, \qquad (6.1)$$

$$\hat{g}_{i;c} = \bar{g}_{i;c} + \delta\hat{g}_{i;c} \,,$$

where $\delta\hat{f}_c$ and $\delta\hat{g}_{i;c}$ are the corrections to be added to the trial solutions. Moreover, one iteration of the scoring system is sufficient to yield estimators $\hat{f}_c$ and $\hat{g}_{i;c}$ which are fully efficient, provided the trial solutions $\bar{f}_c$ and $\bar{g}_{i;c}$ are consistent (8, p. 255). Since the Anderson-Lazarsfeld-Dudman estimators (1) have this property, they are used in the present example as the first approximation (Table 3) to the structure.

TABLE 3

First Approximation [Anderson, (1)] to an Efficient Estimate of the Latent Class Structure

Machine design ability		ξ	
Latent classes (domain of ξ)		1 ("creative")	2 ("non-creative")
Latent marginal probabilities		$\tilde{f}_1 = .516228$	$\tilde{f}_2 = .483772$
Subtests x_i	Domain of x_i	Subtest conditional probabilities	
x_1	1	$\tilde{g}_{1;1} = .729419$	$\tilde{g}_{1;2} = .080928$
	0	$\tilde{g}_{\bar{1};1} = .270581$	$\tilde{g}_{\bar{1};2} = .919072$
x_2	1	$\tilde{g}_{2;1} = .745642$	$\tilde{g}_{2;2} = .188564$
	0	$\tilde{g}_{\bar{2};1} = .254358$	$\tilde{g}_{\bar{2};2} = .811436$
x_3	1	$\tilde{g}_{3;1} = .774096$	$\tilde{g}_{3;2} = .112673$
	0	$\tilde{g}_{\bar{3};1} = .225904$	$\tilde{g}_{\bar{3};2} = .887327$
x_4	1	$\tilde{g}_{4;1} = .741515$	$\tilde{g}_{4;2} = .226393$
	0	$\tilde{g}_{\bar{4};1} = .2584$[illegible]$5$	$\tilde{g}_{\bar{4};2} = .773607$

For the latent class structure postulated to underlie the data of Table 2, an efficient estimate is given in Table 4. As an example of Table 4, $\hat{f}_1 = .581917$ because from Table 3, $\bar{f}_1 = .516228$ and, by the method of calculation explained below, $\delta\hat{f}_1 = .065689$; hence, using (6.1),

$$\hat{f}_1 = \bar{f}_1 + \delta\hat{f}_1 = .516228 + .065689 = .581917.$$

TABLE 4

An Efficient Estimate of the Latent Class Structure

Machine design ability	ξ	
Latent classes (domain of ξ)	1 ("creative")	2("non-creative")
Latent marginal probabilities	$\hat{f}_1 = .581917$	$\hat{f}_2 = .418083$

Subtests x_i	Domain of x_i	Subtest conditional probabilities	
x_1	1	$\hat{g}_{1;1} = .686730$	$\hat{g}_{1;2} = .114614$
	0	$\hat{g}_{\bar{1};1} = .313270$	$\hat{g}_{\bar{1};2} = .885386$
x_2	1	$\hat{g}_{2;1} = .728419$	$\hat{g}_{2;2} = .173071$
	0	$\hat{g}_{\bar{2};1} = .271581$	$\hat{g}_{\bar{2};2} = .826929$
x_3	1	$\hat{g}_{3;1} = .676457$	$\hat{g}_{3;2} = .078556$
	0	$\hat{g}_{\bar{3};1} = .323543$	$\hat{g}_{\bar{3};2} = .921444$
x_4	1	$\hat{g}_{4;1} = .676904$	$\hat{g}_{4;2} = .173385$
	0	$\hat{g}_{\bar{4};1} = .323096$	$\hat{g}_{\bar{4};2} = .826615$

Finally, the corrections $\delta \hat{f}_c$ and $\delta \hat{g}_{i;c}$ are obtained by the scoring system from the relations:

$$\delta \hat{f}_c = \frac{1}{n} \sum_{c'=1}^{\gamma-1} n\tilde{I}^{(c)(c')} \tilde{S}_{c'} + \frac{1}{n} \sum_{i=1}^{p} \sum_{c''=1}^{\gamma} n\tilde{I}^{(c)(i;c'')} \tilde{S}_{i;c''}$$

$$(c = 1, \cdots, \gamma - 1); \tag{6.2}$$

$$\delta \hat{g}_{i;c} = \frac{1}{n} \sum_{c'=1}^{\gamma-1} n\tilde{I}^{(c')(i;c)} \tilde{S}_{c'} + \frac{1}{n} \sum_{i''=1}^{p} \sum_{c''=1}^{\gamma} n\tilde{I}^{(i;c)(i'';c'')} \tilde{S}_{i'';c''}$$

$$(i = 1, \cdots, p; c = 1, \cdots, \gamma).$$

Here the tilde on the $\tilde{S}$ and the $\tilde{I}$ indicate evaluation at the trial solutions $\tilde{f}_c$ and $\tilde{g}_{i;c}$ of:

(*i*) the S functions, one for each structural parameter, which are defined as

$$S_c = \frac{\partial(\log L)}{\partial f_c} \qquad (c = 1, \cdots, \gamma - 1),$$

$$S_{i;c} = \frac{\partial(\log L)}{\partial g_{i;c}} \qquad (i = 1, \cdots, p; c = 1, \cdots, \gamma), \tag{6.3}$$

i.e., simply a relabeling of the left side of (4.2). These functions are referred to as the *efficient scores* for the parameters f_c and $g_{i;c}$. For the data of Table 2, the efficient scores as evaluated at the first approximation are given in Table 5.

(*ii*) the inverse matrix of the I, or *information*, functions. Taking the

TABLE 5

Efficient Scores $\widetilde{S}_c$ and $\widetilde{S}_{i;c}$ Evaluated at the First Approximation

$\widetilde{S}_1 = 10.635378$	
$\widetilde{S}_{1;1} = 3.624402$	$\widetilde{S}_{1;2} = 35.114865$
$\widetilde{S}_{2;1} = 1.260259$	$\widetilde{S}_{2;2} = 6.861212$
$\widetilde{S}_{3;1} = -25.120091$	$\widetilde{S}_{3;2} = -5.782032$
$\widetilde{S}_{4;1} = -16.637863$	$\widetilde{S}_{4;2} = -9.284267$

TABLE 6(a)

Information Matrix per Single Observation, Evaluated at the First Approximation

		f_1	$g_{1;1}$	...	$g_{3;2}$	$g_{4;2}$
1	f_1	3.048439	0.509498	...	0.678498	0.365129
2	$g_{1;1}$	0.509498	2.017739	...	-0.526254	-0.255081
3	$g_{2;1}$	0.354050	-0.124733	...	-0.392939	-0.178048
.	.	.	.		.	.
.	.		.		.	.
.	.	.	.		.	.
9	$g_{4;2}$	0.365129	-0.255081	...	-0.127786	2.136945

TABLE 6(b)

Inverse of Information Matrix per Single Observation, Evaluated at the First Approximation

		f_1	$g_{1;1}$	...	$g_{3;2}$	$g_{4;2}$
1	f_1	0.588484	0.254609	...	-0.211503	-0.183881
2	$g_{1;1}$	-0.254609	0.677139	...	0.202876	0.157569
3	$g_{2;1}$	-0.187074	0.118647	...	0.141973	0.109468
.	.	.	.		.	.
.	.		.		.	.
.	.	.	.		.	.
9	$g_{4;2}$	-0.183881	0.157569	...	0.093239	0.561545

parameters pairwise, the information functions, expressed per single observation, are:

$$\begin{aligned}
\frac{1}{n} I_{(c)(c')} &= \frac{1}{n} E(S_c S_{c'}) \qquad (c, c' = 1, \cdots, \gamma - 1), \\
\frac{1}{n} I_{(i;c)(i'';c'')} &= \frac{1}{n} E(S_{i;c} S_{i'';c''}) \\
&\qquad (i, i'' = 1, \cdots, p; c, c'' = 1, \cdots, \gamma), \\
\frac{1}{n} I_{(c')(i;c)} &= \frac{1}{n} E(S_{c'} S_{i;c}) \\
&\qquad (c' = 1, \cdots, \gamma - 1; i = 1, \cdots, p; c = 1, \cdots, \gamma),
\end{aligned} \tag{6.4}$$

i.e., essentially the variances and covariances of the efficient scores. The square matrix of these elements, of order $\gamma - 1 + p\gamma$, is denoted by I and is called the *information* matrix. Hence in (6.2), $n\tilde{I}^{(c)(c')}$, $n\tilde{I}^{(c')(i;c)}$, and $n\tilde{I}^{(i;c)(i'';c'')}$ constitute the elements of the inverse of the information matrix, after evaluation at the trial solutions. For the present machine design data, the matrix $(1/n)\tilde{I}$ is given in Table 6(a) and $n\tilde{I}^{-1}$ in Table 6(b). Explicit formulas for the calculation of (6.3) and (6.4) are given in the Appendix.

As an example of the use of (6.2),

$$\delta\hat{f}_1 = \frac{1}{n}\,[n\tilde{I}^{(1)(1)}\tilde{S}_1 + n\tilde{I}^{(1)(1;1)}\tilde{S}_{1;1} + n\tilde{I}^{(1)(2;1)}\tilde{S}_{2;1} + \cdots + n\tilde{I}^{(1)(4;2)}\tilde{S}_{4;2}]$$

$$= \frac{1}{137}\,[(.588484)(10.635378) + (-.254609)(3.624402)$$

$$+ (-.187074)(1.260259) + \cdots + (-.183881)(-9.284267) = .065689,$$

using Table 5 and Table 6(b).

From (6.2) the crucial step in the scoring system is the inversion of the information matrix, i.e., the existence of I^{-1} is a necessary condition for identifiability. That the non-singular character of I is also a sufficient condition has been noted in section 3.

Since I^{-1} is the asymptotic variance-covariance matrix of the efficient estimators of the structural parameters, it is often desired to obtain the first approximation inverse matrix $\tilde{I}^{-1}$ explicitly. [As an example of the use of $\tilde{I}^{-1}$, from Table 6(b), the variance of $\hat{f}_1$ is approximately $\tilde{V}(\hat{f}_1) = .588484/137 = .00429$]. However, it is worth noting that if estimates of the large sample variances and covariances are not sought, then (6.2) is equivalent to solving a set of simultaneous linear equations in the corrections $\delta\hat{f}_c$ and $\delta\hat{\theta}_{i;c}$ without explicitly inverting the matrix $\tilde{I}$. That is, as a matrix equation (6.2) is $\hat{\delta} = \tilde{I}^{-1}\tilde{S}$, which is equivalent to $\tilde{I}\hat{\delta} = \tilde{S}$, where

$$\hat{\delta} = \begin{bmatrix} \delta\hat{f}_1 \\ \vdots \\ \delta\hat{\theta}_{p;\gamma} \end{bmatrix} \quad \text{and} \quad \tilde{S} = \begin{bmatrix} \tilde{S}_1 \\ \vdots \\ \tilde{S}_{p;\gamma} \end{bmatrix}.$$

For the machine design data, (5.1) is applied to Table 2 in order to test for the number of significant latent classes, yielding

$$\chi^2 = (23 - 18.276211)^2/18.276211 + \cdots$$

$$+ (34 - 32.650525)^2/32.650525 = 33.01$$

with $2^4 - 2(4 + 1) = 6$ degrees of freedom. Since the probability of a chi square this large by pure chance is less than .001, there may well be justifica-

tion, with respect to the universe of examinees and psychological test used (12), for rejecting the hypothesis of $\gamma = 2$ latent classes for the latent class model and for assuming the existence of more than two classes.

Appendix

Evaluation of the $\tilde{S}$ and $\tilde{I}$ Functions for Tables 5 *and* 6(a).

(*i*) *The efficient scores $\tilde{S}$ of Table* 5.

Differentiation of the logarithm of (4.1) and evaluation at the first approximation leads to the general computation formulas

$$\tilde{S}_c = \sum_D \tilde{q}_{(c)(i_1 i_2 \cdots i_p)} n_{i_1 i_2 \cdots i_p} ,$$
$$\tilde{S}_{i;c} = \sum_D \tilde{q}_{(i;c)(i_1 i_2 \cdots i_p)} n_{i_1 i_2 \cdots i_p} . \tag{1}$$

Thus, the efficient scores are linear functions of the response frequencies $n_{i_1 i_2 \cdots i_p}$, with coefficients $\tilde{q}$ which are functions of $\tilde{f}_c$ and $\tilde{g}_{i;c}$, viz., the quotients:

$$\tilde{q}_{(c)(i_1 i_2 \cdots i_p)} = \frac{c\text{th term of } (\tilde{g}_{i_1 i_2 \cdots i_p}/\tilde{f}_c)}{\tilde{g}_{i_1 i_2 \cdots i_p}} ,$$
$$\tilde{q}_{(i;c)(i_1 i_2 \cdots i_p)} = \frac{c\text{th term of } (\tilde{g}_{i_1 i_2 \cdots i_p}/\tilde{g}_{i;c})}{\tilde{g}_{i_1 i_2 \cdots i_p}} . \tag{2}$$

That is, the denominator of a coefficient $\tilde{q}$ is that manifest marginal $g_{i_1 i_2 \cdots i_p}$ indicated by the second subscript on $\tilde{q}$, and evaluated at the trial solution. In order to obtain the numerator of $\tilde{q}$ readily, it is convenient in the determination of the denominator to compute each of the γ terms separately (later adding them, of course, to get the complete denominator). Then as (2) shows, the numerator is easily obtained in one operation, viz., by dividing the cth term of the denominator by $\tilde{f}_c$ (if dealing with $\tilde{S}_c$) or by $\tilde{g}_{i;c}$ (if dealing with $\tilde{S}_{i;c}$).

For example, from (1), for $p = 4$ and $\gamma = 2$,

$$\tilde{S}_{3;2} = \tilde{q}_{(3;2)(1234)} n_{1234} + \tilde{q}_{(3;2)(\bar{1}234)} n_{\bar{1}234} + \cdots + \tilde{q}_{(3;2)(\bar{1}\bar{2}\bar{3}\bar{4})} n_{\bar{1}\bar{2}\bar{3}\bar{4}} ,$$

which, for the data of section 6, is $= -5.782032$ as shown in Table 5. For the first coefficient, $\tilde{q}_{(3;2)(1234)}$, the manifest marginal for the denominator is g_{1234}, corresponding to the second subscript 1234 on $\tilde{q}$, which from (2.2) is

$$g_{1234} = f_1 g_{1;1} g_{2;1} g_{3;1} g_{4;1} + f_2 g_{1;2} g_{2;2} g_{3;2} g_{4;2} .$$

Hence, the denominator is obtained as the sum of the following two parts ($c = 1, 2 = \gamma$): $\tilde{f}_1 \tilde{g}_{1;1} \tilde{g}_{2;1} \tilde{g}_{3;1} \tilde{g}_{4;1}$ and $\tilde{f}_2 \tilde{g}_{1;2} \tilde{g}_{2;2} \tilde{g}_{3;2} \tilde{g}_{4;2}$. The numerator of $\tilde{q}_{(3;2)(1234)}$ is obtained by dividing $\tilde{g}_{3;2}$ into the $c = 2$nd term of the denomina-

tor, $\tilde{f}_2\tilde{g}_{1:2}\tilde{g}_{2:2}\tilde{g}_{3:2}\tilde{g}_{4:2}$, giving $\tilde{f}_2\tilde{g}_{1:2}\tilde{g}_{2:2}\tilde{g}_{4:2}$. Hence by (2) the first coefficient is

$$\tilde{q}_{(3:2)(1234)} = \frac{\tilde{f}_2\tilde{g}_{1:2}\tilde{g}_{2:2}\tilde{g}_{4:2}}{\tilde{f}_1\tilde{g}_{1:1}\tilde{g}_{2:1}\tilde{g}_{3:1}\tilde{g}_{4:1} + \tilde{f}_2\tilde{g}_{1:2}\tilde{g}_{2:2}\tilde{g}_{3:2}\tilde{g}_{4:2}},$$

which can be calculated by direct substitution from Table 3 to be .010358. The other coefficients are obtained in a similar manner.

Actually, for a given efficient score $\tilde{S}$, only half of the numerators of the coefficients $\tilde{q}$ need be calculated. For example, out of 16 numerators for $\tilde{S}_{3:2}$ only 8 of the numerators need be computed, each of these being the negative of one of the remaining 8, e.g., $\tilde{f}_2\tilde{g}_{1:2}\tilde{g}_{2:2}\tilde{g}_{4:2}$, which occurs in $\tilde{q}_{(3:2)(1234)}$, is the negative of $-\tilde{f}_2\tilde{g}_{1:2}\tilde{g}_{2:2}\tilde{g}_{4:2}$, which occurs in $\tilde{q}_{(3:2)(12\bar{3}\bar{4})}$.

(*ii*) *The information functions,* $\tilde{I}$, *of Table* 6(a).

Evaluation of the expectations (6.4) at the first approximation leads to the general computational formulas:

$$\begin{aligned}
\frac{1}{n}\tilde{I}_{(c)(c')} &= \sum_D \frac{[\text{numerator of } \tilde{q}_{(c)(i_1 i_2 \cdots i_p)}][\text{numerator of } \tilde{q}_{(c')(i_1 i_2 \cdots i_p)}]}{\tilde{g}_{i_1 i_2 \cdots i_p}}, \\
\frac{1}{n}\tilde{I}_{(i:c)(i':c')} &= \sum_D \frac{[\text{numerator of } \tilde{q}_{(i:c)(i_1 i_2 \cdots i_p)}][\text{numerator of } \tilde{q}_{(i':c')(i_1 i_2 \cdots i_p)}]}{\tilde{g}_{i_1 i_2 \cdots i_p}}, \\
\frac{1}{n}\tilde{I}_{(c')(i:c)} &= \sum_D \frac{[\text{numerator of } \tilde{q}_{(c')(i_1 i_2 \cdots i_p)}][\text{numerator of } \tilde{q}_{(i:c)(i_1 i_2 \cdots i_p)}]}{\tilde{g}_{i_1 i_2 \cdots i_p}}.
\end{aligned} \tag{3}$$

Thus, comparing (3) with (2), one notes that Table 6(a) can be obtained from the calculations already made in completing Table 5.

For example, from (3) and (2), for $p = 4$ and $\gamma = 2$,

$$\frac{1}{n}\tilde{I}_{(2:1)(3:2)} = \frac{[(\text{1st term of } \tilde{g}_{1234})/\tilde{g}_{2:1}][(\text{2nd term of } \tilde{g}_{1234})/\tilde{g}_{3:2}]}{\tilde{g}_{1234}}$$

$$+ \cdots + \frac{[(\text{1st term of } \tilde{g}_{\bar{1}\bar{2}\bar{3}\bar{4}})/\tilde{g}_{2:1}][(\text{2nd term of } \tilde{g}_{\bar{1}\bar{2}\bar{3}\bar{4}})/\tilde{g}_{3:2}]}{\tilde{g}_{\bar{1}\bar{2}\bar{3}\bar{4}}},$$

which for the data of section 6 is $-.392939$. Specifically, the first of these terms for $(1/n)\tilde{I}_{(2:1)(3:2)}$ is

$$\frac{[\tilde{f}_1\tilde{g}_{1:1}\tilde{g}_{3:1}\tilde{g}_{4:1}][\tilde{f}_2\tilde{g}_{1:2}\tilde{g}_{2:2}\tilde{g}_{4:2}]}{\tilde{f}_1\tilde{g}_{1:1}\tilde{g}_{2:1}\tilde{g}_{3:1}\tilde{g}_{4:1} + \tilde{f}_2\tilde{g}_{1:2}\tilde{g}_{2:2}\tilde{g}_{3:2}\tilde{g}_{4:2}},$$

which can be computed directly from the calculations previously made for $\tilde{q}_{(3:2)(1234)}$ of $\tilde{S}_{3:2}$ [as illustrated in (*i*) above] and for $\tilde{q}_{(2:1)(1234)}$ of $\tilde{S}_{2:1}$, which can be similarly illustrated.

REFERENCES

1. Anderson, T. W. On estimation of parameters in latent structure analysis. *Psychometrika*, 1954, **19**, 1-10.
2. Coombs, C. H. A theory of psychological scaling. Eng. Res. Bull. No. 34, University of Michigan, 1952.
3. Fisher, R. A. The conditions under which chi square measures the discrepancy between observation and hypothesis. *J. R. stat. Soc.*, 1924, **87**, 442-450.
4. Gulliksen, H. Theory of mental tests. New York; Wiley, 1950.
5. Hurwicz, L. Generalization of the concept of identification. Statistical inference in dynamic economic models. Cowles Commission Monograph 10. 1950, 245-257.
6. Koopmans, T. C. and Reiersøl, O. The identification of structural characteristics. *Ann. math. Stat.*, 1950, **21**, 165-181.
7. Lord, F. A theory of test scores. Psychometric Monograph, No. 7, 1952.
8. Neyman, J. Contribution to the theory of the χ^2-test. Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University California Press, 1949.
9. Neyman, J. and Scott, E. L. Consistent estimates based on partially consistent observations. *Econometrica*, 1948, **16**, 1-32.
10. Rao, C. R. Advanced statistical methods in biometric research. New York: Wiley, 1952.
11. Reiersøl, O. On the identifiability of parameters in Thurstone's multiple-factor analysis. *Psychometrika*, 1950, **15**, 121-149.
12. Schumacher, C. F., Maxson, G. R., and Martinek, H. Tests for creative ability in machine design. Annual report, 1953. Project ONR 458. Armed Services Technical Information Agency 21 284.
13. Stevens, S. S. Handbook of experimental psychology. New York: Wiley, 1951.
14. Stouffer, S. A., Lazarsfeld, P. F., et al. Measurement and prediction. Princeton: Princeton Univ. Press, 1950.

Manuscript received 5/12/55

Revised manuscript received 9/15/55

PSYCHOMETRIKA—VOL. 21, NO. 4
DECEMBER, 1956

THE SELECTION OF JUDGES FOR PREFERENCE TESTING*

R. DARRELL BOCK

UNIVERSITY OF CHICAGO

A scheme for choosing a few individuals whose preferences for given objects are most representative of those of a larger group of individuals is proposed. The method involves (*a*) quantifying the preferences of each individual so as to discriminate optimally among objects, (*b*) testing statistically whether or not a common preference continuum may be assumed for the quantified preferences, (*c*) constructing a linear estimator of values for the objects on this continuum, if it may be assumed, and (*d*) selecting as judges the least number of individuals whose quantified preferences, when used with this estimator, determine values for the objects with acceptable accuracy. A numerical example based on food preferences is presented.

I. *Introduction*

Preference studies, particularly of foods, frequently depend upon a limited number of judges who have been chosen from a larger group of individuals. Common practice is to choose those individuals whose repeated preferences for the same objects are most reliable. This assumes that all individuals reflect a common dimension of taste and that an individual whose preferences are reliable, but atypical, will not be encountered. A more thorough method would include a test for dimensionality and, if this condition is met, the selection of judges who are most representative of the group as a whole.

Let us assume that the preference of the ith individual for the jth object on the kth occasion is determined by a preference score g_{ijk}, which has the composition

$$g_{ijk} = \alpha_i \omega_j + e_{ijk}, \tag{1}$$

where α_i is a coefficient characterizing the ith individual, ω_j is a value characterizing the jth object, and e_{ijk} is a random error distributed as $N(0, \zeta_i^2)$ and independently of all α_i, ω_j, and other e_{ijk}. As usual, the variance of g_{ijk} over all occasions and objects is unity and the variance of ω_j over all objects is taken so that

$$\alpha_i^2 + \zeta_i^2 = 1. \tag{2}$$

*Preparation of this paper has been supported in part by the Quartermaster Food and Container Institute for the Armed Forces. Views or conclusions expressed herein are those of the author and do not necessarily reflect the views or indorsement of the Department of Defense.

The problem is to test the compatability of the observed preferences with (1) and to derive a scheme for estimating the ω_j which will most efficiently utilize a limited number of judges.

The preference scores are not observable, of course, but are presumed to underlie ratings, rankings, or paired comparisons of the objects, and are assumed to be estimable only by resort to a scaling method which quantifies the preferences. In the present context, an efficient and practical method is to choose scale values for the preferences which will yield a minimal estimate of ζ_i^2. This type of scaling has been proposed by Fisher (6), in a different connection by Guttman (7, 8), and by Maung (16), Johnson (12), and Bartlett (4). Related tests of significance have been discussed by Bartlett (2), Fisher (5, 6), Williams (19), Marriott (15), and others. A suitable version of the method for the present purpose is as follows.

II. *Scaling the Preferences*

$$
\begin{array}{cc|cccc|c}
 & & & \text{Objects} & & & \\
 & & 1 & 2 & \cdots & q & \\
\hline
 & 1 & t_{11} & t_{12} & \cdots & t_{1q} & \\
 & 2 & t_{21} & t_{22} & \cdots & t_{2q} & \\
\text{Occasions} & \vdots & \vdots & \vdots & t_{kj} & \vdots & \\
 & p & t_{p1} & t_{p2} & \cdots & t_{pq} & \\
\hline\hline
 & 1 & n_{.11} & n_{.21} & \cdots & n_{.q1} & n_{..1} \\
 & 2 & n_{.12} & n_{.22} & \cdots & n_{.q2} & n_{..2} \\
t & \vdots & \vdots & \vdots & & \vdots & \vdots \\
 & r & n_{.1r} & n_{.2r} & \cdots & n_{.qr} & n_{..r} \\
\hline
 & & n_{.1.} & n_{.2.} & \cdots & n_{.q.} & n_{...}
\end{array}
\tag{3}
$$

t_{kj} designates a category of preference on the kth occasion for the jth object ($t = 1, 2, \cdots, r$):

(a) For ratings, t indicates a quasi-quantitative category such as good, fair, or poor.
(b) For rankings, t indicates a position in an ordering.
(c) For paired comparisons, t indicates the number of times one object is preferred to all others.

$n_{.jt}$ is the frequency with which the tth category occurs for the jth object over the p occasions.

$n_{..t}$ is the frequency of the tth category for all objects.

$$n_{.j.} = p,$$

$$n_{...} = qp = n,$$

$$r \leq q.$$

Representing the estimated score of the tth category by x_t, the average preference score for the jth object on p repetitions is defined by

$$g_{j.} = \frac{\sum_t x_t n_{.jt}}{\sum_t x_{.jt}} = \frac{\sum_t x_t n_{.jt}}{p}, \tag{4}$$

where a subscript for the ith individual is tacitly understood.

Since interest in preference testing is primarily in distinguishing quantitatively between objects, scores should be assigned to the categories which will discriminate optimally among the objects, where optimum discrimination is attained when the sum of squares between objects is a maximum with respect to that within objects. But for sums of squares:

$$\text{Between} + \text{Within} = \text{Total},$$

so this is equivalent to maximizing the between sum of squares subject to the condition that the total sum of squares remains constant. That is, assign values to x_t which maximize

$$\frac{1}{p} \sum_j \left(\sum_t x_t n_{.jt}\right)^2 - \frac{1}{n} \left(\sum_t x_t n_{..t}\right)^2, \tag{5}$$

subject to the condition that

$$\sum_t x_t^2 n_{..t} - \left(\sum_t x_t n_{..t}\right)^2 / n \tag{6}$$

be a finite constant.

Since this maximum is independent of the origin and scale of the x_t, impose the conditions that scores for the sample sum to zero and have unit variance. That is,

$$\sum_t x_t n_{..t} = 0, \tag{7}$$

and

$$\sum_t x_t^2 n_{..t} = n \quad . \tag{8}$$

Accordingly, (5) and (6) become

$$\sum_j \left(\sum_t x_t n_{.jt}\right)^2 / p, \tag{9}$$

and

$$\sum_t x_t^2 n_{..t} \,. \tag{10}$$

To express the required maximum explicitly as a function of the observed frequencies, it is convenient to adopt matrix notation and to define:

$x = [x_1\ x_2 \cdots x_t \cdots x_r]$ = row vector of assigned scores;

$F = [n_{.jt}] = r \times q$ matrix of frequencies in (3);

$FF'/p = H$;

$D = [n_{..t}] = r \times r$ diagonal matrix of marginal frequencies in (3).

Then (9) becomes

$$\sum_j \left(\sum_t x_t n_{.jt}\right)^2/p = xHx', \tag{11}$$

and (10) becomes

$$\sum_t x_t^2 n_{..t} = xDx'. \tag{12}$$

The maximum of xHx', subject to the condition that $xDx' = n$, is obtained by differentiating and equating the resulting expression to zero:

$$xHx' - \eta^2(xDx' - n),$$

where η^2 is a Lagrangian multiplier. The result is

$$x(H - \eta^2 D) = 0, \tag{13}$$

which may be written

$$x(HD^{-1} - \eta^2 I) = 0. \tag{14}$$

For x to be non-null it is necessary that the determinant

$$|\, HD^{-1} - \eta^2 I \,| = 0, \tag{15}$$

where η^2 is a root of (15) and x is the corresponding latent vector of HD^{-1}.

An iterative method [cf. Hartree (10, p. 178ff.)] may be used to obtain latent vectors z_h of HD^{-1} with arbitrary origin and unit. The first of these vectors, with $\eta_1^2 = 1$ and $z_1 = (1\ 1 \cdots 1)$, is trivial and can be avoided by replacing F with the corresponding matrix of deviations from expectation. The second latent vector provides the required maximum of $xHD^{-1}x'$, and its associated root η_2^2 is, as the notation suggests, a correlation ratio giving the fraction of sum of squares between objects. For, noting that a linear transformation of a latent vector is a latent vector, from (13)

$$z_2 H - \eta_2^2 z_2 D = 0.$$

Transposing and post-multiplying by z_2',

$$z_2 H z_2' = \eta_2^2 z_2 D z_2' = \eta_2^2 n. \tag{16}$$

Similarly $\hat{\zeta}^2$, which estimates ζ^2, is the fraction of sum of squares within objects, and

$$\eta_2^2 + \hat{\zeta}^2 = 1. \tag{17}$$

To transform z_2 into x, impose conditions (7) and (8) to obtain

$$x_i = u(z_i - v), \tag{18}$$

where

$$v = \sum_i z_i n_{..i}/n, \tag{19}$$

and

$$u = \frac{\sqrt{n}}{\sqrt{\sum_i (z - v)^2 n_{..i}}}. \tag{20}$$

In general, HD^{-1} has $r - 1$ non-trivial latent vectors, of which the first, x_2, provides estimates of the maximally discriminating scores for the preference categories. In the present context further interpretation for any remaining vector, even if its associated root could be shown significant, does not seem necessary, although it should be noted that relevant interpretations are available (**9**).

The statistical significance of the scores may be questioned on two counts: If they are to be useful for discriminating among the objects, variance attributable to differences in column means (mean preference scores) must be significant. Fisher (**6**) suggests that an approximate test, based on an analysis of variance, be made directly from the value of η_2^2 as shown in Table 1.

TABLE 1

Form of Analysis of Variance for Derived Scores

Source of variation	Degrees of freedom	Sums of squares	Mean squares	F
Between objects	$q-1+r-1$	$\eta_2^2 n$	$\dfrac{\eta_2^2 n}{q+r-2}$	$\dfrac{\eta_2^2 q(p-1)-(r-1)}{(1-\eta_2^2)(q+r-2)}$
Residual	$q(p-1)-(r-1)$	$n-\eta_2^2 n$	$\dfrac{n(1-\eta_2^2)}{q(p-1)-(r-1)}$	
Total	$n-1$	n		

[Cf. Williams (19).] The degrees of freedom are adjusted for the arbitrary constants fitted by adding $r - 1$ to those between objects and subtracting $r - 1$ from those of the residual. The scores may be considered useful only if F is significant.

It is also of interest to test whether these derived, maximally discriminating scores are any real improvement on arbitrarily assigned scores, e.g., $1, 2, 3, \cdots, r$. Again following Fisher (6), this may be done by testing the significance of variance between objects, computed from the derived scores, which remains after variance attributable to the assigned scores has been removed. The appropriate analysis of covariance is shown in Table 2. (The

TABLE 2

Form of Analysis of Covariance of Assigned and Derived Scores

Source of variation	Sums of squares (assigned scores)	Sums of cross-products	Sums of squares (derived scores)
Between objects	$\xi H \xi'$	$\xi H x'$	$x H x'$
Residual	$\xi D \xi' - \xi H \xi'$	$\xi D x' - \xi H x'$	$x D x' - x H x'$
Total	$\xi D \xi'$	$\xi D x'$	$x D x'$

vector of arbitrarily assigned scores, must satisfy the relation $\sum_i \xi_i n_{..i} = 0$.)

The analysis of variance of the derived scores, eliminating the assigned scores, is shown in Table 3. The degrees of freedom between objects has been reduced by 1, since the elimination of ξ leaves only $r - 2$ scores adjustable.

TABLE 3

Form of Analysis of Variance of Derived Scores, Eliminating Assigned Scores

Source of variation	Degrees of freedom	Sums of squares	Mean squares	F
Between objects	$q+r-3$	SS_b : Obtained by subtraction	$\frac{SS_b}{q+r-3}$	$\frac{SS_b q(p-1)-(r-1)}{SS_r(q+r-3)}$
Residual	$q(p-1)$ $-(r-1)$	$SS_r = \{xDx' - xHx'\}$ $-\left\{\frac{(\xi Dx' - \xi Hx')^2}{\xi D\xi' - \xi H\xi'}\right\}$	$\frac{SS_r}{q(p-1)-(r-1)}$	
Total	$n-2$	$SS_t = xDx' - \frac{(\xi Dx')^2}{\xi D\xi'}$		

A significant value of F is necessary, of course, if the derived scores are to be preferred to more convenient assigned scores. As before, this test is only approximate. For a more exact treatment, see Bartlett (4).

III. *Estimating the ω_j*

If written explicitly, the scaled preferences of N individuals for g objects on p occasions would appear as in Table 4. For subsequent computations

TABLE 4

Preference Scores and Mean Preference Scores

	Objects															
	1				2				···	q						
	Individuals				Individuals				···	Individuals						
Occasions	1	2	···	N	1	2	···	N	···	1	2	···	N			
1	g_{111}	g_{211}	···	g_{N11}	g_{121}	g_{221}	···	g_{N21}	···	g_{1q1}	g_{2q1}	···	g_{Nq1}			
2	g_{112}	g_{212}	···	g_{N12}	g_{122}	g_{222}	···	g_{N22}	···	g_{1q2}	g_{2q2}	···	g_{Nq2}			
⋮	⋮	⋮		⋮	⋮	⋮		⋮		⋮	⋮		⋮			
p	g_{11p}	g_{21p}	···	g_{N1p}	g_{12p}	g_{22p}	···	g_{N2p}	···	g_{1qp}	g_{2qp}	···	g_{Nqp}			
Means	$g_{11.}$	$g_{21.}$	···	$g_{N1.}$	$g_{12.}$	$g_{22.}$	···	$g_{N2.}$	···	$g_{1q.}$	$g_{2q.}$	···	$g_{Nq.}$			

only the mean preference scores are needed and may be obtained by the relation

$$g_{ij.} = x_i n_{ij}/p, \tag{21}$$

where x_i is the first non-trivial latent vector for the ith individual under the transformation (18), and n_{ij} is the jth column of the corresponding F matrix.

A method of estimating the ω_j from the preference scores of the N individuals is needed. Resorting again to the requirement of optimal discrimination among the objects, choose as an estimator a linear compound of the preference scores which maximizes the sum of squares of the combined scores between objects on condition that the total sum of squares of the combined scores is constant. Let the estimate of ω_j on the kth occasion be

$$w_{jk} = \sum_i y_i g_{ijk}, \tag{22}$$

and for all occasions

$$w_{j.} = \sum_i y_i g_{ij.}. \tag{23}$$

The analysis of variance of the preference scores with respect to this linear compound may be represented as:

	SS	DF
Between Objects	yMy'	$q - 1$
Within Objects	yBy'	$q(p - 1)$
Total	yTy'	$qp - 1$

where M, B, and T are, respectively, matrices of between objects, within objects, and total sums of squares and cross products for the preference scores as represented in Table 4. Only M must be computed explicitly, with

$$M = pGG', \tag{24}$$

where G is an $N \times q$ matrix of mean preference scores $[g_{ij.}]$.

The coefficients of the required optimally discriminating linear compound are the elements of y determined from

$$y(M - \lambda T) = 0, \tag{25}$$

where λ is the largest root of

$$| M - \lambda T | = 0. \tag{26}$$

If the preferences are made independently, and the objects are identical on successive occasions, the assumption that the preferences of different individuals for the same object are uncorrelated over occasions appears justified. In this case the off-diagonal elements of T and M are asymptotically equal and

$$B = T - M$$

approximates an $N \times N$ diagonal matrix with elements

$$b_{ii} = \sum_j \sum_k g_{ijk}^2 - p \sum_j g_{ij.}^2 .$$

By (8) and (16) this becomes

$$\begin{aligned} b_{ii} &= n - (px_iF_iF_i'x_i')/p^2 \\ &= n - n\eta_i^2 , \end{aligned}$$

and by (17)

$$b_{ii} = n\zeta_i^2 . \tag{27}$$

Taking $Z = [n\hat{\zeta}_i^2]$ as an approximation of B,

$$| M - \lambda(M + Z) | = 0,$$

or, letting

$$\mu = m\lambda/(1 - \lambda), \tag{28}$$

with $m = q(p - 1) - (r - 1)$, the additional $r - 1$ degrees of freedom being removed because of the constants fitted in scaling,

$$| MZ^{-1} - \mu I/m | = 0. \tag{29}$$

Taking the largest root of (29), y may be determined from

$$y(MZ^{-1} - \mu I/m) = 0. \tag{30}$$

[Cf. Bartlett (3).]

If μ_1 is clearly significant, a basis for testing of the significance of variation attributable to the remaining roots is provided by Bartlett's (2) approximation for Wilks' ratio

$$\chi^2 = -[n - \tfrac{1}{2}(N + q)] \sum_{s=1}^{\min(N,q-1)} \log_e (1 - \mu_s). \tag{31}$$

For large n this becomes approximately

$$\chi^2 = \sum_{s=1}^{\min(N,q-1)} \mu_s . \tag{32}$$

That is, the latent vectors associated with each of the canonical variances transform MZ^{-1} into $\min(N, q - 1)$ asymptotically independent quadratic forms, each distributed as χ^2 with degrees of freedom equal to the number of arbitrary constants fitted. Specifically, the sum of the canonical variances,

$$m \cdot \mathrm{tr}\,(MZ^{-1}) = \mu_1 + \mu_2 + \cdots + \mu_{\min(N,q-1)} , \tag{33}$$

is distributed as the sum of $\min(N, q - 1)$ asymptotically independent chi squares with degrees of freedom

$$N(q - 1) = (N + q - 2) + (N + q - 4) + \cdots + \{N - q - 2[\min (N, q - 1)]\}.$$

To test the residual variance, eliminate that of the first root by taking

$$\chi^2 = m \cdot \mathrm{tr}\,(MZ^{-1}) - \mu_1 \tag{34}$$

on $N(q - 1) - (N + q - 2)$ degrees of freedom.

Equation (29) has in general $\min(N, q - 1)$ roots, $\mu_1 , \mu_2 , \cdots , \mu_s$, when ordered by size, giving the canonical variances of MZ^{-1}. If (1) holds, only μ_1 will be significant and the linear compound specified by the corresponding latent vector y_1 will account for all the significant variation in the preference scores. In critical cases (31) would be somewhat more accurate.

It is of interest to note that if (1) holds and p is large,

$$g_{ijs} = \alpha_i \omega_j$$

approximately, and

$$G = \alpha'\omega.$$

If

$$\omega\omega'/q = 1,$$

$$GG'/q = \alpha'\alpha = M/n.$$

Then (30) becomes approximately

$$y(n\alpha'\alpha - \mu B/n) = 0,$$

where

$$| n\alpha'\alpha - \mu B/n | = 0$$

has only one root:

$$\mu = n^2\alpha B^{-1}\alpha', \tag{35}$$

and one associated latent vector:

$$\begin{aligned} y &= n\alpha B^{-1} \\ &\cong n\alpha Z^{-1}, \end{aligned} \tag{36}$$

a result identical in form with that obtained when constructing a linear discriminant function for two groups [cf. Kendall (14, vol. II, p. 341)] and similar also to that for the estimation of mental factors [cf. Holzinger and Harman (11, p. 322)]. Furthermore, from (2) and (17)

$$\alpha_i^2 = \eta_i^2 ,$$

and, asymptotically,

$$Z = [n\zeta_i^2],$$

or, by (36)

$$y_i = \eta_i/(1 - \eta_i^2). \tag{37}$$

Hence, when unidimensionality of tastes can be assumed, the optimal weights for combining preferences are strictly analogous to those for combining tests scores on the basis of reliability [cf. Kelley (13, p. 211)]. For, if α is considered the correlation coefficient between a fallible score and a theoretical "true" score, then

$$\alpha_i^2 = r_{i\infty}r_{i\infty} = r_{ii} = \eta_i^2 ,$$

and

$$y_i = \sqrt{r_{ii}}/(1 - r_{ii}).$$

IV. *Selecting the Judges*

In the sample the best estimator of ω, is given by the first latent vector y associated with the solution of (30). Individuals for whom the corresponding elements of y are larger are to be preferred as judges. If k such individuals

are selected, the question arises whether the loss of information caused by excluding the preferences of the remaining $N - k$ individuals from the estimates of the ω_i is appreciable.

If y_k is the vector y after the elements for the unwanted individuals have been dropped, and M_k and Z_k^{-1} are M and Z^{-1} with corresponding $N - k$ rows and columns dropped, then adjusting y_k so that $y_k y_k' = 1$, the portion of the canonical variance μ_1 which is attributable to the k selected individuals is

$$v_k = m y_k M_k Z_k^{-1} y_k' . \tag{38}$$

The significance of the remaining variance can be tested approximately by taking

$$\mu_1 - v_k \tag{39}$$

as χ^2 with $(N - k)(q - 1)$ degrees of freedom [cf. Rao (**18**, p. 257)]. If this variance is insignificant, or appropriately small with respect to μ_1, the preferences of the k selected judges can be considered adequate for estimating the ω_i.

Since the preferences within objects are uncorrelated, the elements remaining in y_k are unchanged except in scale. If it is convenient to have standarized values for the w_{ik} within objects and between occasions, the elements of y_k should be multiplied by

$$\left[(\hat{\zeta}_1^2 y_1^2 + \hat{\zeta}_2^2 y_2^2 + \cdots + \hat{\zeta}_k^2 y_k^2) \Big/ m \right]^{-1/2} . \tag{40}$$

V. *Related Problems*

It should be understood that the sample dealt with here is, in effect, a series of occasions out of infinitely many in which the preferences of the same individuals might be repeated under the same conditions. Predictions made on the basis of this sample apply only to future preferences of these individuals and are not immediately generalizable to the population from which the individuals are drawn. When a single significant dimension of taste is found, however, this limitation is less serious because there is no evidence which contradicts the assumption that a new group from the population will share the same dimension of taste and show the same relative preferences for the objects. When significant additional dimensions are found, the problem becomes much less tractable but in some ways more interesting. Additional dimensions raise the question of what and how many attributes in the objects are being evaluated by the individuals and reflected in the preferences. This is essentially a problem in multidimensional psychophysics; the type of canonical analysis used in the present study appears to be an alternative to approaching the problem by a direct reduction of a table of proportions from paired comparisons of the objects. Preference scores for the additional

dimensions may be estimated using the other significant latent vectors from (30), and rotation of these vectors to yield more meaningful scores is permissible. In practical applications, determining dimensionality could be important because it would indicate the minimum number of attributes which must be controlled if the objects are to be produced in uniform quality. A natural extension of this would be to identify the attributes physically or chemically and subject them to planned control and development.

The converse problem would be to identify and characterize by independent variables those individuals whose preferences are responsible for or most representative of the various dimensions of taste. If individuals with atypical tastes were few, they could be excluded from further preference testing and the remaining homogeneous group considered representative of the ultimate consumers of the objects. Alternatively, it might be found that distinct subgroups of the individuals with different tastes can be identified in terms of sex, race, income-group, etc. A set of separate panels representing each of these subgroups would, so to speak, "span the space of preferences" in the population. If the proportion of these groups in the population were known, the reaction of the total population to new objects could be predicted by a weighted combination of the preferences of the subgroups. Finally, the relationships among the preferences of the groups could be studied by a canonical analysis or a multivariate analysis of variance in which the "individuals" of this study are replaced by groups and "occasions" by individuals within groups. A problem of this sort based on a national food preference survey is now under study.

VI. *Numerical Example*

Oberman and Li (17) report ten replicate ratings (p) of ten individuals (N) for pastries baked from five different fats (q). Ratings were made on a five-point scale (r): 1 = not edible; 2 = poor; 3 = fair; 4 = good; 5 = excellent. For the first individual, the ratings of the five pastries are shown in Table 5.

The body of Table 5 is the matrix F; the frequencies on the right are the elements of the diagonal matrix D. Then:

$$pH = FF' = \begin{bmatrix} 62 & 42 & 35 & 01 & 00 \\ 42 & 60 & 34 & 04 & 00 \\ 35 & 34 & 38 & 10 & 03 \\ 01 & 04 & 10 & 37 & 18 \\ 00 & 00 & 03 & 18 & 09 \end{bmatrix}$$

$$\text{Column sum:} \quad 140 \quad 140 \quad 120 \quad 70 \quad 30$$

The rows and columns of pH must sum to $pn_{..t}$. To form HD^{-1}, divide elements in the tth column of pH by $pn_{..t}$. It is convenient at the same time to remove the expectation by subtracting $1/r$ from each of the resulting elements. Designate the resulting matrix by T:

$$T = \begin{bmatrix} .2429 & .1000 & .0917 & -.1857 & -.2000 \\ .1000 & .2286 & .0833 & -.1429 & -.2000 \\ .0500 & .0428 & .1167 & -.0571 & -.1000 \\ -.1929 & -.1714 & -.1167 & .3286 & .4000 \\ -.2000 & -.2000 & -.1750 & .0571 & .1000 \end{bmatrix}$$

Column sum:	.0000	.0000	.0000	.0000	.0000

Columns of T must sum to zero. [Note: χ^2 for F, with $(q-1)(r-1)$ degrees of freedom is $n(\text{tr } T)$.]

The first latent vector of T is z_2, the first non-trivial latent vector of HD^{-1}. To extract z_2 from T, premultiply by a trial vector orthogonal to the trivial vector (1 1 1 1 1). The vector (1 1 0 −1 −1) is convenient. For these data, five iterations produce the following essentially accurate estimate:

$$z_2 \cong (.8325 \quad .7831 \quad .6021 \quad -.7721 \quad -1.000),$$

with associated root $\eta_2^2 = .7921$. Using (18), (19), and (20), impose conditions (7) and (8) to obtain

$$(.6292 \quad .5514 \quad .2694 \quad -1.8691 \quad -2.2260).$$

Analysis of variance based on η_2^2 shows significant discrimination between fats (Table 6). Analysis of covariance of assigned scores 1, 2, 3, 4, 5 [adjusted to $\xi = (1.42\ 0.42\ -.58\ -1.58\ -2.58)$ so that $\sum_t \xi_t n_{..t} = 0$] and the derived scores is shown in Table 7. Analysis of variance of the derived scores, eliminating the assigned scores, shows that the derived scores discriminate between fats significantly better than the assigned scores (Table 8).

Scales for the remaining individuals are constructed in the same way. The correlation ratios and results of the significance tests for each individual are shown in Table 9. The mean preference scores of each individual for the pastries are shown in Table 10, the body of which comprises the matrix G. (Rows of G must sum to zero.) The elements of the diagonal matrix Z computed by (27) and (17) from the correlation ratios of Table 9 are:

10.395, 32.690, 32.345, 34.885, 19.355, 13.795, 31.625, 3.365, 24.845, 9.265.

The product $MZ^{-1} = pGG'Z^{-1}$ is given in Table 11.

The first latent vector of MZ^{-1} is extracted by premultiplying by $(1 \cdots 1)$

TABLE 5

Preferences of the First Individual

Categories (t)	Fats 1	2	3	4	5	Total
5	4	3	6	1	0	14
4	2	6	2	4	0	14
3	4	1	2	4	1	12
2	0	0	0	1	6	7
1	0	0	0	0	3	3
Total	10	10	10	10	10	50

TABLE 6

Analysis of Variance for Derived Scores

Source of variation	Degrees of freedom	Sums of squares	Mean squares	F
Between fats	8	39.60	4.95	19.49
Residual	41	10.40	.254	
Total	49	50.00		$p < .01$

TABLE 7

Analysis of Covariance of Assigned and Derived Scores

Source of variation	Sums of squares (assigned scores)	Sums of cross-products	Sums of squares (derived scores)
Between fats	44.08	41.01	39.60
Residual	28.10	10.77	10.39
Total	72.18	51.78	49.99

TABLE 8

Analysis of Variance of the Derived Scores, Eliminating the Assigned Scores

Source of variation	Degrees of freedom	Sums of squares	Mean squares	F
Between fats	7	6.58	.940	6.14
Residual	41	6.26	.153	
Total	48	12.84		$p < .01$

TABLE 9

Scaling Information for All Individuals

Individuals (i)	η_i^2	Significance of discrimination between fats	Significance of variation of derived scores, eliminating unit scores
1	.7921	$p < .01$	$p < .01$
2	.3462	$p < .01$	$p > .05$
3	.3531	$p < .01$	$p \geq .05$
4	.3023	$.05 > p > .01$	$p > .05$
5	.6129	$p < .01$	$.05 > p > .01$
6	.7241	$p < .01$	$p < .01$
7	.3675	$p < .01$	$p > .05$
8	.9327	$p < .01$	$p < .01$
9	.5031	$p < .01$	$p > .05$
10	.8147	$p < .01$	$p < .01$

TABLE 10

Mean Preference Scores of All Individuals for All Fats

Individuals	Fats				
	1	2	3	4	5
1	.4700	.5460	.5420	.2041	-1.7622
2	.5845	.2507	-.0919	.3502	-1.0934
3	.3678	.2466	.4890	.0485	-1.1522
4	.4828	.0496	.6780	-.8101	-.4005
5	.5975	.4179	.6315	-.1997	-1.4472
6	.5832	.5512	.5279	-.0199	-1.6422
7	.4460	.4534	.4311	-.2402	-1.0905
8	.9541	.8092	.5789	-1.0702	-1.2723
9	.5092	.5456	.6024	-.4948	-1.1623
10	.5065	.5988	.4568	.2265	-1.7884
Total	5.5016	4:4692	4.8460	-2.0055	-12.8113

TABLE 11

The Matrix MZ^{-1}

$$\begin{bmatrix}
3.8095 & .7219 & .8079 & .3331 & 1.7365 & 2.7191 & .8108 & 9.5929 & 1.1314 & 4.3287 \\
2.2703 & .5295 & .4665 & .1108 & .9861 & 1.6086 & .4563 & 5.1233 & .5943 & 2.6325 \\
2.5137 & .4616 & .5458 & .2705 & 1.1830 & 1.8123 & .5477 & 6.6805 & .7774 & 2.8376 \\
1.1178 & .1182 & .2917 & .4334 & .7639 & .9717 & .3673 & 6.7459 & .6229 & 1.2051 \\
3.2333 & .5838 & .7078 & .4238 & 1.5833 & 2.3867 & .7445 & 9.8930 & 1.0841 & 3.6530 \\
3.6085 & .6788 & .7729 & .3843 & 1.7011 & 2.6238 & .8011 & 10.1605 & 1.1407 & 4.1004 \\
2.4666 & .4414 & .5355 & .3330 & 1.2165 & 1.8365 & .5810 & 7.9851 & .8537 & 2.7960 \\
3.1053 & .5274 & .6950 & .6507 & 1.7200 & 2.4784 & .8496 & 13.8618 & 1.3220 & 3.5246 \\
2.7042 & .4517 & .5972 & .4436 & 1.3916 & 2.0544 & .6707 & 9.7608 & 1.0125 & 3.0507 \\
3.8581 & .7461 & .8128 & .3201 & 1.7486 & 2.7539 & .8191 & 9.7043 & 1.1377 & 4.3967
\end{bmatrix}$$

TABLE 12

χ^2 Test for the Dimensionality of G

Source of χ^2	Degrees of freedom	χ^2	P
First canonical variance	13	1066.3239	
Residual	27	138.1454	$p < .01$
Total	40	1204.4693	

as the first trial vector. Four iterations yield an essentially stable first latent root $(1/m)\ \mu = 26.0079$ and first latent vector (with arbitrary unit):

$$(.2885 \quad .0520 \quad .0630 \quad .0426 \quad .1455 \quad .2179 \quad .0699 \quad 1.0000 \quad .1038 \quad .3276).$$

The χ^2 test of variance remaining after that attributable to the first root has been removed shows that the preference scores *cannot* be considered unidimensional (Table 12).

Inspection of the mean preference scores (Table 10) suggests that the departure from unidimensionality is a result of disagreement between the scores of the most reliable individual, no. 8, and those of the remaining more reliable individuals, nos. 1, 5, 6, and 10. Reliability is judged from the correlation ratios in Table 9. Individuals whose preferences show low reliability influence the test of dimensionality only slightly compared with those of high reliability. The disagreement is most marked for pastries baked from fat no. 4.

When results of this sort are encountered in practice, a careful review of the conditions of testing, the stability of the objects, and the training of the individuals is probably indicated. In the present example it would be particularly unfortunate to exclude individual no. 8, who is highly reliable, if through better controlled testing, retraining, etc., his preferences could be brought into line with those of other judges.

On the other hand, a single dimension in this example accounts for $26.0079/29.3773 = 88.53\%$ of the variance of the preference scores; hence, the error incurred by ignoring the residual variance might not be considered important in practice. In this case, individuals 1, 5, 6, 8, and 10 could be accepted as the limited panel of judges without further attention to differences among their preferences. The loss of information resulting from the exclusion of the remaining individuals is indicated by the difference in variance given by (39):

$$41(26.0079 - 23.4348) = 105.4971.$$

Taking this difference as χ^2 with 20 degrees of freedom indicates that the loss of variance is significant. It is clear, however, that the marginal increase in accuracy gained from any additional individual probably would not merit the added expense or inconvenience.

Adjusting the scale of the y_i for the selected individuals according to (40) yields

$$w_{jk} = .7720g_{1jk} + .3893g_{5jk} + .5831g_{6jk} + 2.6758g_{8jk} + .8766g_{10jk}$$

as the optimal linear estimator of the ω_{jk} with unit variance within objects and between occasions.

It is of interest to compare the coefficients of this estimator with those

derived from the correlation ratios for these individuals by prior assumption of unidimensional preference scores. The resulting coefficients, brought to the same scale, are

$$.7922, \quad .3743, \quad .5708, \quad 2.6558, \quad .9015.$$

The close agreement of corresponding coefficients reflects the large proportion of variance in the preference scores accounted for by the first canonical variance.

The values for the pastries baked from the five fats are determined by applying the optimal linear estimator to the mean preference scores:

	Fat				
	1	2	3	4	5
Value:	3.933	3.596	2.922	−2.597	−7.854

These values must sum to zero.

The use of the facilities of the Agricultural Experimental Station of the University of Puerto Rico for preparing the numerical example is gratefully acknowledged.

REFERENCES

1. Bartlett, M. S. Further aspects of the theory of multiple regression. *Proc. Camb. phil. Soc.*, 1938, **34**, 33-40.
2. Bartlett, M. S. Multivariate analysis. *J. roy. statist. Soc., Suppl.*, 1947, **9**, 176-190.
3. Bartlett, M. S. Internal and external factor analysis. *Brit. J. Psychol., Statist. Sect.*, 1948, **1**, 73-81.
4. Bartlett, M. S. The goodness of fit of a single hypothetical discriminant function in the case of several groups. *Ann. Eugen.*, 1951, **16**, 199-214.
5. Fisher, R. A. The precision of discriminant functions, *Ann. Eugen.*, 1940, **10**, 422-429.
6. Fisher, R. A. Statistical methods for research workers. New York: Hafner, 1948.
7. Guttman, L. The quantification of a class of attributes: a theory and method of scale construction. In P. Horst et al., The prediction of personal adjustment. New York: Social Science Research Council, 1941, pp. 319-348.
8. Guttman, L. An approach for quantifying paired comparisons and rank order. *Ann. math. Statist.*, 1946, **17**, 144-163.
9. Guttman, L. Components of scale analysis. In Stouffer et al., Measurement and prediction. Princeton: Princeton Univ. Press, 1950, pp. 312-361.
10. Hartree, D. R. Numerical analysis. Oxford: Clarendon Press, 1952.
11. Holzinger, K. J. and Harman, H. H. Factor analysis. Chicago: Univ. Chicago Press, 1941.
12. Johnson, P. O. The quantification of data in discriminant analysis. *J. A. S. A.*, 1950, **45**, 65-76.
13. Kelley, T. L. Interpretation of educational measurements. Yonkers-on-Hudson: World Book, 1927.
14. Kendall, M. J. The advanced theory of statistics. London: Griffin, 1948.

15. Marriott, F. H. C. Tests of significance in canonical analysis. *Biometrika*, 1952, **39**, 58-64.
16. Maung, Khint. Discriminant analysis of Tocher's eye-color data for Scottish school children. *Ann. Eugen.*, 1941, **11**, 64-76.
17. Overman, A. and Li, J. C. R. Dependability of food judges as indicated by analysis of scores of food testing panel. *Food Res.*, 1948, **13**, 441-449.
18. Rao, C. R. Advanced statistical methods in biometric research. New York: Wiley, 1952.
19. Williams, E. J. Use of scores for the analysis of association in contingency tables. *Biometrika*, 1952, **39**, 274-289.

Manuscript received 5/28/55

Revised manuscript received 10/24/55

PSYCHOMETRIKA—VOL. 21, NO. 4
DECEMBER, 1956

AN EMPIRICAL EVALUATION OF MULTIDIMENSIONAL SUCCESSIVE INTERVALS*

SAMUEL J. MESSICK†

UNIVERSITY OF ILLINOIS

The multidimensional method of successive intervals and the method of complete triads are applied to similarity judgments of Munsell colors varying in brightness, saturation, and hue. Both methods yield configurations that correlate highly with the Munsell color structure. This validation of these scaling methods in an area of known dimensionality indicates their applicability for exploration in areas of unknown dimensionality.

Several recent multidimensional scaling procedures (**2**, **7**, **14**, **18**), based upon a Euclidean mathematical model (**19**), reveal the number and the nature of relevant dimensions in unknown areas. Empirical evaluations of the methods in areas of known dimensionality seem appropriate before applying them in in areas of unknown dimensionality. The field of color perception was selected for such validation, since the dimensions of brightness, saturation, and hue are well defined.

Richardson (**14**) applied the method of triadic combinations to judgments of similarity among some Munsell colors, which differed in saturation and brightness but were of a constant hue; the results are reported as being in essential agreement with the Munsell scheme. Torgerson (**17**) applied the complete method of triads to judgments of similarity among nine Munsell colors, which were presumably of the same red hue but differed in brightness and saturation. His two-dimensional configuration was very similar to the Munsell system. Torgerson also compared complete triads with unidimensional paired comparisons using nine gray stimuli that differed only in brightness. The methods yielded unidimensional scales that were linearly related. These results can be looked upon as a "validation" of this multidimensional scaling method. Indications are that saturation and brightness, at least, can be represented adequately by a Euclidean model (**18**).

Since the method of complete triads requires so many judgments, the task becomes prohibitive with more than ten stimuli. In order to overcome this difficulty, a multidimensional method of successive intervals (**1**, **8**) was

*This study was supported in part by the Office of Naval Research contract N6onr-270-20 with Princeton University and also in part by funds from the National Science Foundation. The research was carried out when the author was an Educational Testing Service Psychometric Fellow at Princeton University.

†Now at The Menninger Foundation.

recently developed; this method requires fewer judgments per subject and thus permits the use of a larger number of stimuli. This is an important consideration in investigations of unknown areas, for although the dimensionality does not necessarily increase with the number of stimuli, it is certainly limited by that number.

This investigation is an evaluation of the multidimensional method of successive intervals in the area of color perception. With the new procedure it is feasible to include wide stimulus variations, so the study is also a step toward a systematic multidimensional mapping of psychological color space.

If the stimuli are chosen from the Munsell color system (**11**), a comparison between the Munsell scale values and those obtained from the multidimensional procedure will permit an evaluation of the method. Since Torgerson has already applied a multidimensional scaling procedure (complete triads) to color stimuli with acceptable results (**17**), a comparison of multidimensional successive intervals with complete triads can also be looked upon as a validation procedure. One aspect of this comparison will be an investigation of the difference in difficulty of the judgments required. In complete triads, which presumably involves a much simpler judgment, the subject must decide which two of three stimuli are most similar. Successive intervals requires the subject to decide whether the members of pair A are more or less alike than the members of pair B and then to order the pairs accordingly on a continuum. The practical importance of this difference may be evaluated not only in terms of the similarity of the final structures but also in terms of the time required to make the judgments, the ease with which they are made, and their acceptability to the subject as reasonable tasks.

The Methods

The multidimensional scaling methods are techniques for estimating the interpoint distances among a set of stimuli. [For detailed discussion of these methods, see (**9**).] By requiring judgments about pairs of stimuli instead of single stimuli, these methods yield scale values which can be taken to represent the distances between the members of the pairs (**20**). The general procedure is to obtain similarity judgments among stimuli and then to scale these judgments by traditional scaling methods (**5, 15, 16**). The scale values obtained represent distances between stimuli, which can be analyzed to obtain the dimensionality of the space and the projections of the stimuli on a set of axes placed in the space (**10**).

In complete triads each combination of three stimuli (triad) is presented three times; subjects are asked on separate occasions whether A is more like B or C, whether B is more like A or C, and whether C is more like A or B. Since each of $n(n-1)(n-2)/6$ possible triads is presented three times, $n(n-1)(n-2)/2$ judgments are required from each subject. In the multidimensional method of successive intervals (**1, 8**), on the other hand, the

subject is asked only to arrange the $n(n - 1)/2$ possible pairs of n stimuli into $(k + 1)$ categories on a distance continuum according to the degree of similarity of the members of each pair. This procedure is a direct extension of unidimensional successive intervals (4, 6), with pairs of stimuli substituted for single stimuli.

Experimental Procedures

The Stimuli and Their Mode of Presentation

Eight of the nine stimuli used by Torgerson were scaled by the method of complete triads. According to the Munsell system (11), these stimuli are of the same red hue but differ in brightness and saturation. Eight additional Munsell colors of the same characteristics were scaled by multidimensional successive intervals. Although these colors are of the same hue according to the Munsell system, the report of the Optical Society of America subcommittee on the spacing of the Munsell colors (13) indicates slight hue variations. Since the *OSA* renotations represent an attempt to obtain a closer approximation to equal-appearing intervals than the spacing in the Munsell scheme (12), they would seem to be superior as criteria.

The *OSA* revised designations of the 16 stimuli scaled by multidimensional successive intervals are plotted in Figure 1. The circled values are the

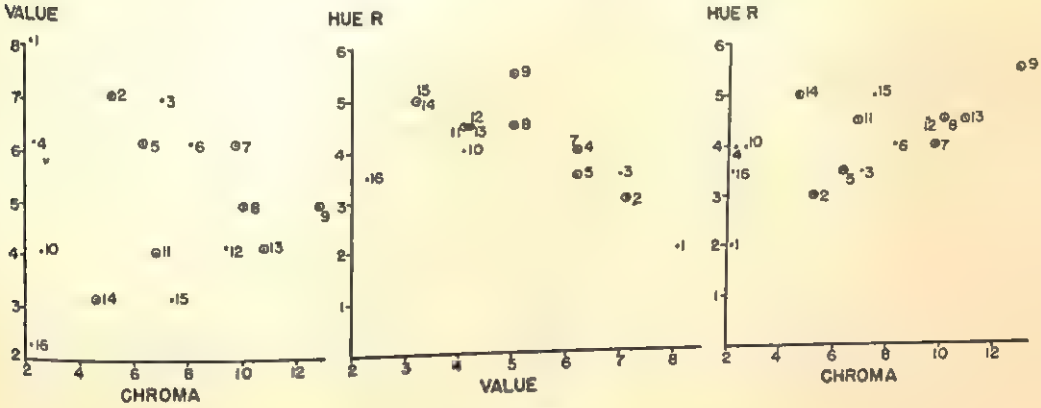


FIGURE 1
Stimulus Configuration According to OSA Revisions

eight stimuli also scaled by the method of complete triads. The Munsell notation designates the three dimensions of color perception as hue, value (brightness), and chroma (saturation). Theoretically steps along the value and chroma scales in the Munsell system represent equal sense distances along the respective psychological dimensions, two chroma steps representing a sense distance approximately equal to one value step.

For the method of complete triads, sheets of colored paper obtained

from the Munsell Color Company were cut into equilateral triangles 1.5 inches on a side. Each of the 56 combinations of three different colored triangles was then mounted on a white cardboard triangle 6.25 inches on a side. For the multidimensional method of successive intervals, the colored sheets were cut into one-inch squares, and each pair of different colored squares (120 pairs for 16 stimuli) was mounted on a white 3″ × 5″ card. Each stimulus appeared as often on the left as on the right. The cards for both procedures were randomly arranged for presentation to the subjects. The stimuli were viewed against the white background of the mounts, which in turn were presented against a gray background. The lighting source was a ceiling fixture containing two GE 40-watt fluorescent daylight bulbs.

The Subjects

Forty-two subjects, 38 males and four females, took part. Color-blind persons were excluded by tests. The subjects were randomly divided into two equal groups, members of one group judging complete triads first and members of the other group judging successive intervals first.

Instructions

For the complete triads task, the subjects were to decide whether the top color on a triangular card looked more like the left one or the right one. The response of "right" or "left" was recorded by the experimenter. The subjects were also told that the colors on some triangles were very much alike, but that no triangle had colors exactly the same. They were encouraged to respond with a first impression and not to spend too much time on any one triad.

For the multidimensional successive intervals task, the subjects were to sort the stack of rectangular cards into eight piles. First they were to divide it into two approximately equal piles according to whether the colors on each card were very similar or very different. Then the subjects were to divide each of those piles into two again. If at any time a subject wished to change a card from one pile to another, he was permitted to do so. Then the subjects were asked to divide each of the four piles into two in the same way, giving eight piles. The left-hand piles were to contain cards on which the colors were the most similar, and the right-hand pile those on which the colors were the most different. In going from left to right the cards should get increasingly different. In the final step, the subjects looked through each pile to see if the cards in that pile seemed to "go together" and to make any necessary changes.

The Number of Judgments Required

The number of judgments for multidimensional successive intervals is minimal if each subject looks at each stimulus card only once and then

decides in which pile it belongs. The minimum number of judgments, then, is equal to the number of stimulus cards [$n(n - 1)/2$]. Use of this procedure would have required 120 judgments from each subject in order to scale 16 stimuli. It was desired, however, to have the subject re-sort the cards after he had become familiar with them; using an expanded multidimensional successive intervals procedure, such check sorts would be obtained without the number of judgments approaching appreciably that required by complete triads. Since each of the $n(n - 1)/2$ cards was looked at four times, $2n(n - 1)$ or 480 judgments were made by each subject in order to scale 16 stimuli. If these 16 stimuli had been scaled by complete triads, however, $n(n - 1)(n - 2)/2$ or 1680 judgments would have been required.

In the complete triads section of the present experiment, each of the 56 triads was presented three times, making a total of 168 judgments required from each subject to scale eight stimuli.

Analysis and Results

The Method of Complete Triads

The raw data for the complete triads section of the study consisted of a 42 (subjects) × 168 (triads) table of responses. These data were analyzed by the complete triads procedure (**18**) to obtain a matrix of relative interstimulus distances. The general multidimensional scaling procedure (**10**) was then applied to obtain a matrix of projections, which is analogous to a factor matrix. Its rank, r, is equal to the dimensionality of the Euclidean space defined by the experimentally obtained distances, and its elements represent the projections of the stimuli on a set of r orthogonal axes placed at the centroid of the stimuli.

The matrix of projections was of rank 2. A third dimension might have been expected to represent the differences in hue. Its absence may be accounted for by the fact that the hue variations were slight compared with the range of variations in brightness and saturation. It is also possible that eight variables were too few to distinguish a third factor from error. A detailed description of the analysis and a tabulation of the data appear in (**8**). The two-dimensional structure was rotated orthogonally to approximate the Munsell dimensions. The complete triad structure and the revised Munsell structure agree closely, as can be seen in a comparison of each triads factor with the corresponding Munsell dimension (Figure 2). The present structure also had about the same degree of agreement with Torgerson's results (**17**).

The Multidimensional Method of Successive Intervals

The raw data for the successive intervals method were summarized in a 120 (stimulus-pairs) × 8 (categories) table of the number of times the ith pair of stimuli was placed in the gth category. Some of these stimulus-pairs

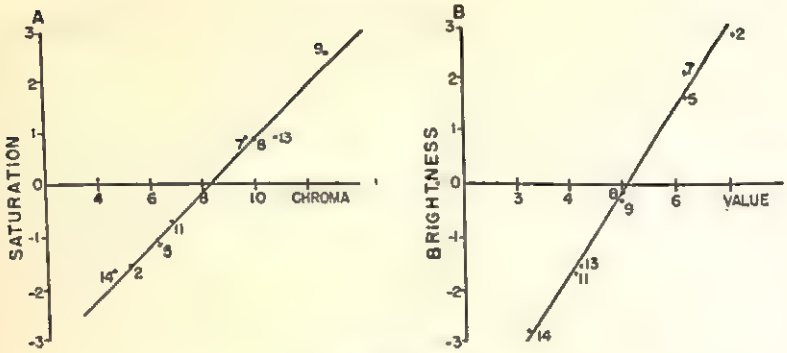


FIGURE 2
Separate Comparisons of Triad Dimensions
vs. Revised Munsell Dimensions

had been unanimously or nearly unanimously placed in one of the extreme categories, making it impossible to obtain a scale value for every pair by ordinary successive intervals procedures. If four stimuli (stimuli 6, 12, 15, and 16) were excluded from the set, however, the frequency distributions of the remaining pairs had satisfactory ranges. Thus, the set of distances among the remaining 12 stimuli was scaled by the method of successive intervals.

By the selection of two other sets of stimuli with satisfactory distributions (a set composed of the ten stimuli 1, 2, 4, 5, 6, 8, 9, 10, 12, and 15 and a set composed of the six stimuli 4, 8, 10, 12, 15, and 16), it was possible to obtain distance estimates involving all 16 stimuli. All possible distances among the 16 stimuli could not be obtained, of course, but it was hoped that there would be sufficient overlap among the three sets to allow the four excluded points to be fitted into the space defined by the set of 12 stimuli. Since some of the distances were involved in all three sets, the similarity of overlapping estimates could be evaluated. If the common distances were similarly estimated in all three groups, it would seem reasonable to attempt to fit the excluded four points into the space defined by the set of 12 stimuli.

When the three sets of stimuli were analyzed separately, the common distances were found to be estimated similarly in all three sets. Accordingly, the four excluded stimuli were fitted into the space of the 12 stimuli by a least squares criterion (8). The results for the set of 12 stimuli only will be considered in this paper, but it is suffice to say that this procedure located the four excluded points in positions reasonably appropriate to the Munsell scheme.

The raw data for the set of 12 stimuli, then, consisted of a 66 (stimulus-pairs) $\times$ 8 (categories) table of the number of times the ith stimulus-pair was placed in the gth category. Successive intervals procedures (3) yielded the inter-stimulus distances, and the general multidimensional scaling procedure (10) was applied to obtain a matrix of projections, which consisted

of three factors. The three dimensions were rotated orthogonally to approximate the Munsell configuration. The Munsell and the successive intervals configurations are correlated to an extremely high degree. The extent of the agreement is seen in a comparison of each successive intervals factor with the

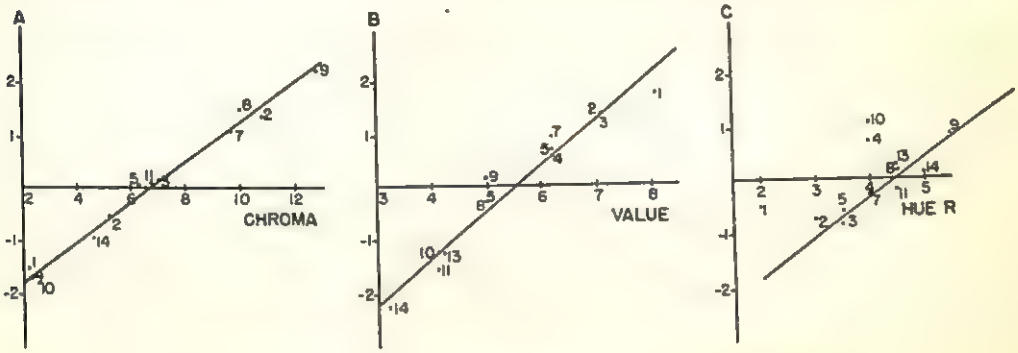


FIGURE 3
Separate Comparisons of Successive Intervals Dimensions vs. Revised Munsell Dimensions

corresponding Munsell dimension (Figure 3). Factor *A* corresponds to saturation, Factor *B* to brightness, and Factor *C* to hue.

These comparisons reveal only three widely deviant values, the hue positions obtained by successive intervals for stimuli 1, 4, and 10, which were the most unsaturated colors used. When such stimuli are compared in close proximity to stimuli with obvious hue, contrast effects might exaggerate relative inter-stimulus distances. It is also possible, however, that these deviant values indicate a slight deviation from orthogonality in the placement of the reference axes. With oblique rotation procedures these three points may be re-aligned to produce a linear plot in Figure 3 for hue vs. Factor *C*. Such an oblique orientation for Factor *C* might be considered an indication that the orthogonal axis placement in the Munsell system is somewhat in error. This is only a very tentative suggestion, but it illustrates that multidimensional scaling procedures, being based upon cross-dimensional judgments, can lead to such statements about relationships between dimensions. On the other hand, these three deviant points in Figure 3 might also be taken to indicate a slight inadequacy on the part of the Euclidean model to describe psychological color data.

A Comparison Between Complete Triads and Multidimensional Successive Intervals

A comparison between the final rotated structures obtained by complete triads and multidimensional successive intervals indicated excellent agreement. The fact that three dimensions were obtained from the successive

intervals procedure and only two from complete triads is probably more a function of the stimuli than of any difference between the two techniques. Only half the number of stimuli in the successive intervals experiment were scaled by complete triads, and the variations in hue involved were slight compared with the saturation and brightness variations.

The average length of time for the 168 judgments by complete triads was 35 minutes, whereas the average time for the 480 judgments in the successive intervals task was approximately 40 minutes. It is evident that the successive intervals judgment, although presumably more difficult than the triads judgment, was made with considerable ease. Every subject accepted the task in a casual manner, without complaining about its difficulty.

Discussion

Since multidimensional scaling procedures yielded structures which correlated highly with the revised Munsell system, it would now seem reasonable to apply these procedures for purposes of exploration and discovery in areas of unknown dimensionality as well as for confirmation and modification in other areas of known dimensionality. The results of this experiment also indicate the desirability of a systematic multidimensional scaling of psychological color space and, perhaps, a modification of existing color scales. The Munsell color system and the *OSA* revised scales are based upon psychophysical investigations of each dimension separately (12), a procedure which does not permit the comparison of color samples across dimensions. Since the multidimensional scaling approach does allow multidimensional comparisons, a systematic application of these techniques in the color domain should permit the construction of a color solid which is based upon the relationships between as well as within dimensions.

REFERENCES

1. Abelson, R. P. A technique and a model for multidimensional attitude scaling. *Amer. Psychologist*, 1954, **9**, 319.
2. Attneave, F. Dimensions of similarity. *Amer. J. Psychol.*, 1950, **63**, 516-556.
3. Diederich, G. W., Messick, S. J., and Tucker, L. R. A general least squares solution for successive intervals. Princeton: Educational Testing Service Research Bulletin, 1955.
4. Edwards, A. L. The scaling of stimuli by the method of successive intervals. *J. Appl. Psychol.*, 1952, **36**, 118-122.
5. Guilford, J. P. The method of paired comparisons as a psychometric method. *Psychol. Rev.*, 1928, **35**, 494-506.
6. Gulliksen, H. A least squares solution for successive intervals assuming unequal standard deviations. *Psychometrika*, 1954, **19**, 117-139.
7. Klingberg, F. L. Studies in measurement of the relations between sovereign states. *Psychometrika*, 1941, **6**, 335-352.
8. Messick, S. J. The perception of attitude relationships: a multidimensional scaling approach to the structuring of social attitudes. Ph.D. thesis, Princeton Univer., 1954. Also Educational Testing Service Research Bulletin, 1954.

9. Messick, S. J. Some recent theoretical developments in multidimensional scaling. *Educ. Psychol. Measmt.*, 1956, **16**, 82-100.
10. Messick, S. J. and Abelson, R. P. The additive constant problem in multidimensional scaling. *Psychometrika*, 1956, **21**, 1-16.
11. Munsell Book of Color, Abridged Edition. Baltimore: Munsell Color Co., Inc., 1945.
12. Newhall, S. M. The ratio method in the review of the Munsell colors. *Amer. J. Psychol.*, 1939, **52**, 394.
13. Newhall, S. M., Nickerson, D., and Judd, D. B. Final report of the O.S.A. subcommittee on the spacing of the Munsell colors. *J. opt. Soc. Amer.*, 1943, **33**, 385-418.
14. Richardson, M. W. Multidimensional psychophysics. *Psychol. Bull.*, 1938, **35**, 659-660.
15. Saffir, M. A comparative study of scales constructed by three psychophysical methods. *Psychometrika*, 1937, **2**, 179-198.
16. Thurstone, L. L. Rank-order as a psychophysical method. *J. exp. Psychol.*, 1931, **14**, 187-201.
17. Torgerson, W. S. A theoretical and empirical investigation of multidimensional scaling. Ph.D. thesis, Princeton Univer., 1951.
18. Torgerson, W. S. Multidimensional scaling: I. Theory and method. *Psychometrika*, 1952, **17**, 401-419.
19. Young, G. and Householder, A. S. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 1938, **3**, 19-22.
20. Young, G. and Householder, A. S. A note on multidimensional psychophysical analysis. *Psychometrika*, 1941, **6**, 331-333.

Manuscript received 9/1/55

Revised manuscript received 11/14/55

PSYCHOMETRIKA—VOL. 21, NO. 4
DECEMBER, 1956

A NEW SCALING TECHNIQUE FOR ABSOLUTE JUDGMENTS

BURTON S. ROSNER

WEST HAVEN V. A. HOSPITAL AND YALE UNIVERSITY

The results of an experiment using the method of absolute judgments can be viewed as a matrix of conditional probabilities in which the rows represent stimuli and the columns responses. The cosine of the angle between two row vectors is a measure of the similarity of the corresponding stimuli. This cosine provides the basis for a method of scaling the stimuli. Unlike the method of paired comparisons, this new technique does not require arbitrary fixing of a unit of measurement. A numerical example is given.

In the method of absolute judgments (2), the experimenter selects a sample of N stimuli S_i $(i = 1, \cdots, N)$. He presents each stimulus singly to the subject; the order of presentation is randomized. The subject has available a sample of M responses R_k $(k = 1, \cdots, M)$. Each time a stimulus appears, the subject must make one and only one response. During the experiment, every stimulus is presented a number of times. Thus, the experimental data can be displayed as a matrix of conditional probabilities $p_i(k)$. This matrix has N rows, each row corresponding to a stimulus, and M columns, each column representing a response. The entry in any cell (ik) gives the conditional probability of R_k given S_i. Since the subject must make one and only one response on each trial, the sum of the probabilities in any row is 1.00.

Each row of the matrix of conditional probabilities is a vector in M-space. This fact suggests using the cosine of the angle between two row vectors as a measure of the similarity between corresponding stimuli. The cosine of the angle ϕ_{ij} between row vectors i and j, s_{ij}, is given by

$$s_{ij} \equiv \cos \phi_{ij} = \frac{\sum_k p_i(k)p_j(k)}{\sqrt{\sum_k p_i(k)^2 \sum_k p_j(k)^2}}. \tag{1}$$

Since each $p_i(k)$ is greater than or equal to zero, s_{ij} will be positive and will lie in the interval 0.00 to 1.00, inclusive. A value of 0.00 signifies that none of the responses made to S_i is ever made to S_j and vice versa. A value of 1.00 signifies that the distribution of responses to S_i is identical with the distribution for S_j. Values between the upper and lower bounds indicate the extent to which the same responses are made to both stimuli.

The measure of similarity between stimuli, s_{ij}, is in part formally identical with the Pearson product moment correlation—the product moment

correlation is also the cosine of the angle between two vectors. In the latter case, the dimensionality of the space containing those vectors equals the number of joint observations in the sample. The reliability of the product moment correlation depends upon the number of joint observations. But the reliability of s_{ij} depends upon two factors: the number of observations available for determining each $p_i(k)$ and the total number of different responses R_k made to S_i and S_j. Responses made to neither stimulus do not help determine s_{ij}. Thus, a sampling distribution for s_{ij} would be very different from that for the Pearson r. Furthermore, s_{ij} assumes positive or zero values only. Therefore, s_{ij} differs from the Pearson r in several critical properties.

Since s_{ij} increases as the stimuli become more "similar" in the sense of evoking similar distributions of responses, there should be an inverse relation between the psychological scale separation of the stimuli and s_{ij}. Drs. Bert F. Green and Robert P. Abelson have pointed out how this relation can be derived. The derivation leads to a scaling technique which does not require arbitrary fixing of a unit of measurement, as is necessary in the method of paired comparisons (4). This new technique naturally has many features in common with previous scaling methods for absolute judgments (1, 2, 3).

Assume that the responses to S_i are normally distributed, with a mean at the actual position of the stimulus. Thus,

$$p_i(k) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{\bar{R}_k - \bar{S}_i}{\sigma_i}\right)^2\right] d\bar{R}_k\,, \tag{2}$$

where $\bar{R}_k$ and $\bar{S}_i$ are the scale values of R_k and S_i, respectively. To obtain $\sum_k p_i(k)p_j(k)$, use a continuous approximation:

$$\begin{aligned}\sum_k p_i(k)p_j(k) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{\bar{R}_k - \bar{S}_i}{\sigma_i}\right)^2\right] \frac{1}{\sqrt{2\pi}\,\sigma_j} \\ &\qquad \cdot \exp\left[-\frac{1}{2}\left(\frac{\bar{R}_k - \bar{S}_j}{\sigma_j}\right)^2\right] d\bar{R}_k \qquad (3)\\ &= \frac{1}{2\pi\sigma_i\sigma_j} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\left(\frac{\bar{R}_k - \bar{S}_i}{\sigma_i}\right)^2 + \left(\frac{\bar{R}_k - \bar{S}_j}{\sigma_j}\right)^2\right]\right\} d\bar{R}_k\,.\end{aligned}$$

This expression is integrated by completing the square in the exponent. For convenience, let

$$A = (1/\sigma_i^2) + (1/\sigma_j^2), \tag{4}$$

$$B = (\bar{S}_i^2/\sigma_i^2) + (\bar{S}_j^2/\sigma_j^2), \tag{5}$$

and

$$D = (\bar{S}_i/\sigma_i^2) + (\bar{S}_j/\sigma_j^2). \tag{6}$$

Substituting equations (4) through (6) in (3) and completing the square,

$$\sum_k p_i(k)p_j(k) = \frac{1}{2\pi\sigma_i\sigma_j}\int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}[(\sqrt{A}\,\bar{R}_k - D/\sqrt{A})^2 + B - (D/\sqrt{A})^2\}\,d\bar{R}_k$$

$$= \frac{1}{2\pi\sigma_i\sigma_j}\exp\{-\tfrac{1}{2}[B - (D/\sqrt{A})^2]\} \cdot \int_{-\infty}^{\infty}\exp\{-\tfrac{1}{2}A(\bar{R}_k - D/A)^2\}\,d\bar{R}_k \tag{7}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_i\sigma_j}\exp\{-\tfrac{1}{2}[B - (D/\sqrt{A})^2]\}\,\frac{1}{\sqrt{A}}.$$

By resubstituting from equations (4) through (6) and simplifying, (7) becomes

$$\sum_k p_i(k)p_j(k) = \frac{1}{\sqrt{2\pi}\,\sqrt{\sigma_i^2 + \sigma_j^2}}\exp\left\{-\frac{1}{2}\frac{(\bar{S}_i - \bar{S}_j)^2}{\sigma_i^2 + \sigma_j^2}\right\}. \tag{8}$$

If $S_i = S_j$, then (8) reduces to

$$\sum_k p_i(k)^2 = \frac{1}{2\sqrt{\pi}\,\sigma_i}. \tag{9}$$

Thus, evaluating the discriminal dispersions directly:

$$\sigma_i = \frac{1}{2\sqrt{\pi}\,\sum_k p_i(k)^2}. \tag{10}$$

Finally, taking logarithms to the base e on both sides of (8) and simplifying the resulting equation using (1) and (9),

$$(\bar{S}_i - \bar{S}_j)^2 = (\sigma_i^2 + \sigma_j^2)[\log 2 + \log\sigma_i + \log\sigma_j - \log(\sigma_i^2 + \sigma_j^2) - 2\log\cos\phi_{ij}]. \tag{11}$$

When all discriminal dispersions are set equal to unity, (11) becomes

$$\bar{S}_i - \bar{S}_j = 2\sqrt{-\log\cos\phi_{ij}}\,. \tag{12}$$

Reduction of (11) to a computing formula yields

$$(S_i - S_j)^2 = \frac{1}{4\pi}\left(\frac{1}{[\sum_k p_i(k)^2]^2} + \frac{1}{[\sum_k p_j(k)^2]^2}\right) \cdot \left[\log 2 - \log\left(\frac{1}{[\sum_k p_i(k)^2]^2} + \frac{1}{[\sum_k p_j(k)^2]^2}\right) - 2\log\sum_k p_i(k)p_j(k)\right]. \tag{13}$$

TABLE 1

Conditional Probabilities for Line Division Experiment

	Responses														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		.03	.33	.63		.01									
2			.03	.90	.06	.01									
3				.02	.35	.55	.09	.01							
4				.03	.24	.63	.06	.04							
5 Stimuli					.01	.15	.18	.59	.07						
6								.16	.53	.28	.03				
7								.43	.49	.07	.01				
8								.03	.19	.30	.48				
9									.11	.21	.61	.07			

TABLE 2

Inter-stimulus Distances ($\bar{S}_i - \bar{S}_j$)

	S_j								
	1	2	3	4	5	6	7	8	9
1		-0.238	-2.238	-1.884					
2	0.238		-1.700	-1.570					
3	2.238	1.700		-0.161	-1.598				
4	1.884	1.570	0.161		-1.432				
S_i 5			1.598	1.432		-1.507	-1.999		
6					1.507		-0.444	-1.908	-2.155
7					1.999	0.444		-1.514	-1.736
8						1.908	1.514		-0.261
9						2.155	1.736	0.261	

TABLE 3

Successive Stimulus Scale Differences and Scale Values for the Stimuli

	d_{12}	d_{23}	d_{34}	d_{45}	d_{56}	d_{67}	d_{78}	d_{89}
1		2.000	0.354					
2	0.238		0.130					
3	0.538	1.700		1.437				
4	0.314	1.409	0.161					
5			0.166	1.432		0.492		
6					1.507		1.464	0.247
7					1.555	0.444		0.222
8						0.394	1.514	
9						0.419	1.475	0.261
Sum	1.090	5.109	0.811	2.869	3.062	1.749	4.453	0.730
M	0.36	1.70	0.20	1.43	1.53	0.44	1.48	0.24

S_i	1	2	3	4	5	6	7	8	9
$\bar{S}_i$	0.00	0.36	2.07	2.27	3.70	5.23	5.67	7.15	7.39

Mean scale separations between stimuli can be obtained by the usual techniques and can be summed to yield scale values for the stimuli.

To illustrate the use of this scaling method, a modification of an experiment done by W. J. McGill was performed. A single subject sat before a 3-foot square black panel. An opal glass plate, 10 inches long by 4 inches wide, was mounted in the middle of the panel. Every 10 seconds, an 8-inch long horizontal line appeared on the plate for 1 second. A vertical marker on each line extended 1/2 inch above and 1/2 inch below the line. The position of the marker varied from trial to trial. The marker could appear at the midpoint of the horizontal line or 1/2, 1, 1-1/2, or 2 inches to either side of the midpoint. The subject registered judgments of the marker's position by depressing and then returning to normal one of fifteen toggle switches. The switches were mounted in a 14 by 2 by 2-inch box resting in front of the panel. Each stimulus was presented 110 times, but the first 30 trials were not used in scaling the stimuli.

Table 1 shows the matrix of conditional probabilities obtained for this experiment. The inter-stimulus distances computed from (13) appear in Table 2. Table 3 shows estimated distances between stimuli S_i and S_{i+1}. These distances were obtained by subtracting from each non-zero entry of Table 2 the corresponding entry in the next column to the right. The sums and means of the columns of Table 3 appear at the bottoms of the columns. Finally, by setting $\bar{S}_1$ at 0.00 and cumulating distances between adjacent stimuli, the scale values shown in the lower half of Table 3 are obtained. As should be expected, these values show that the subject discriminated poorly between adjacent stimuli on the extreme left or extreme right while discriminating the middle stimuli quite well.

REFERENCES

1. Attneave, F. A method of graded dichotomies for the scaling of judgements. *Psychol. Rev.*, 1949, **56**, 334-340.
2. Garner, W. R. and Hake, H. W. The amount of information in absolute judgements. *Psychol. Rev.*, 1951, **58**, 446-459.
3. Saffir, M. A comparative study of scales constructed by three psychophysical methods. *Psychometrika*, 1937, **2**, 179-198.
4. Thurstone, L. L. Stimulus dispersions in the method of constant stimuli. *J. exper. Psychol.*, 1932, **15**, 284-297.

Manuscript received 9/1/55

Revised manuscript received 11/21/55

PSYCHOMETRIKA—VOL. 21, NO. 4
DECEMBER, 1956

DETERMINATION OF THE NUMBER OF INDEPENDENT PARAMETERS OF A SCORE MATRIX FROM THE EXAMINATION OF RANK ORDERS*

JOSEPH F. BENNETT

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Two ordinal consequences are drawn from the linear multiple-factor analysis model. First, the number $R(s, d)$ of distinct ways in which s subjects can be ranked by linear functions of d factors is limited by the recursive expression $R(s, d) = R(s - 1, d) + (s - 1) R(s - 1, d - 1)$. Second, every set S of $d + 2$ subjects can be separated into two subsets S^* and $S - S^*$ such that no linear function of d variables can rank all S^* over all $S - S^*$, and vice versa. When these results are applied to the hypothetical data of Thurstone's "box problem," three independent parameters are found. Relations to Thurstone's suggestion for a non-correlational factor analysis are discussed.

In the introduction to *Multiple-Factor Analysis* (3, p. xiii), Thurstone suggested that:

> ". . . it would probably be . . . profitable to develop non-metric methods of factor analysis. An idea for such a development would be to determine the number of independent parameters of a score matrix by analyzing successive differences in rank order on the assumption that they are monotonic functions of a limited number of parameters. A score would then be regarded as merely an index of rank order, and that is essentially what we are now doing. The raw scores are transmuted into a normal distribution of unit standard deviation, and these transmuted scores are used for the correlations. Instead of dealing with the transmuted scores in this manner, one might deal with the rank orders directly . . . The actual numerical values are of secondary importance in teasing out . . . the underlying order of a new domain. . . . In putting the results to practical use, the problem would return to a metric form, with standardization and norms which call for statistical methods of the conventional kind."

Workers in the area find the suggestion interesting, but in the present state of knowledge about the relation between ordinal and metric measurement it is impossible to tell whether such a non-metric approach would have practical advantages over present correlational methods, or indeed

*The research reported here was supported jointly by the Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology. The manuscript was prepared while Dr. Bennett was a Fellow of the Center for Advanced Study in the Behavioral Sciences, Stanford, California. His untimely death on May 4, 1956, at the age of 29, is regretfully announced.

whether it is even possible in theory. This paper, to forestall any initial misunderstandings, will *not* attempt to answer these questions. Its purpose is to point out two particular ordinal consequences of the factor analysis model which became apparent during the development of a generalized form (1) of the "Unfolding Method" of Coombs (2), to show how these consequences can be applied in factoring hypothetical data, and to provoke interest in the problems which would have to be solved in applying such methods to empirical data.

General Considerations

Suppose that the factor and population matrices are given. Let the factor scores of each subject be taken as the coordinates of a *subject-vector* in a *population space*, the axes of which represent a set of orthogonal factors. (A subject-vector will refer specifically to its terminus; the term *subject-point* would do as well. For example, the subspace generated by two such vectors will be regarded not as the plane spanned by them and including the origin but as the line connecting their termini and, in general, *not* containing the origin.) It is possible to represent a test in such a space as a line through the origin of the space with appropriate direction cosines relative to the axes, such that the orthogonal projections of the subject-vectors on a test-line are the subjects' scores on that test, up to a linear transformation; the order of those projections on the test-line is the order in which that test ranks the subjects. Although a ranking and its exact inverse will not be distinguished in any important way, it is convenient to give the test-line an orientation, that is, a direction in which scores are to be called *higher*.

Figure 1 represents a two-dimensional population space containing two subject-vectors, A and B. What distinguishes the class of tests which rank A over B from those which rank B over A? Construct a hyperplane, in this case a line called $H(A, B)$, through the origin and normal to the line connecting A and B. It is evident that the two subject-vectors will have coincident projections on $H(A, B)$ and distinct projections on any other line. $H(A, B)$ divides the space into two halfspaces, labeled I and II in Figure 1, such that any test-line whose orientation is from II into I will yield the ranking $A \geq B$. Such a test-line is regarded as "into I." Any test-line into II will rank $B \geq A$. (Such hyperplanes are analogous to Coombs' midpoints; the condition that they shall pass through the origin restricts all rankings to monotonicity.) A test-line coincident with $H(A, B)$ will yield the result $A = B$. In the arguments which follow the convention is adopted that such an assignment of identical scores to two subjects is not a *ranking* but the simultaneous satisfaction of two rankings, $A \geq B$ and $B \geq A$. The distinction will become important in referring to the number of distinct rankings. In Figure 1 there are two rankings, $A \geq B$ and $B \geq A$, but three possible consequences of measurement, $A > B$, $B > A$, and $A = B$.

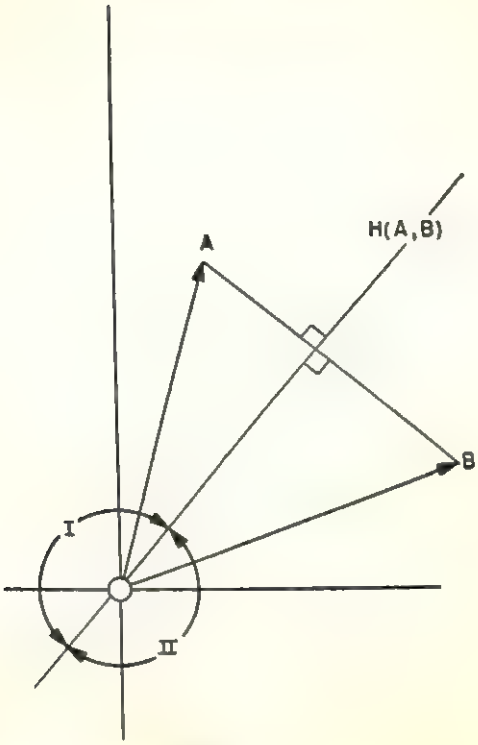


FIGURE 1

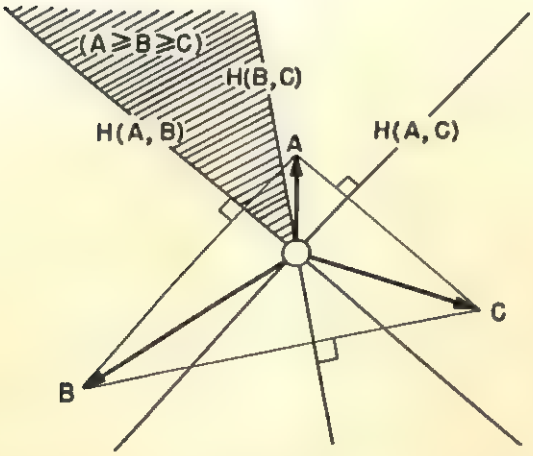


FIGURE 2

Next consider the three subject-vectors in Figure 2. Every pair of subject-vectors will generate a line through the origin which is the unique line on which their orthogonal projections coincide. The three lines so formed will together partition the space into six regions. For example, any test-line into the shaded region shown will rank A over B and B over C. These lines

wholly determine the ordinal properties of the space, because the regions which they form correspond to all the possible rankings which a test could impose on the subjects.

As a final example, consider the four subject-vectors of Figure 3. There

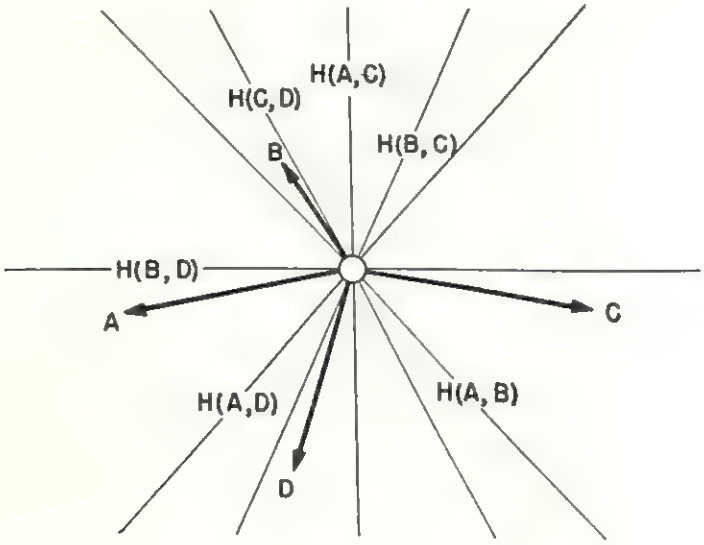


FIGURE 3

is again one line-of-coincident-projection for each pair of subject-vectors, or six such lines in all. Together they divide the space into twelve regions, each region having the property that any test into it will rank all the subjects in the same order. This means that when exactly two factors are present, there are only twelve different ways in which linear functions of those two factors can rank four subjects, and six of these rankings are exact inverses of the other six. (Note that regions which lie opposite to each other across the origin correspond to inverse rankings.) Now algebraically there are twenty-four different ways to rank four subjects; in this construction half of them are missing. This raises two questions with which to begin our investigation: why twelve rankings exactly, rather than ten or fourteen, and why these twelve?

The Number of Linear Rankings

The first question is essentially: how many different ways can s subjects be ranked by linear functions of d variables? Briefly, the answer is as follows: suppose this number is already known for some given number of dimensions and subjects, and it is asked what happens if one more subject is added. One new hyperplane will be formed between each subject already present and the new subject, and each of these hyperplanes will create as many new

regions as it transects, since it cuts each of them into two. The problem is, therefore, to count the number of regions which each hyperplane transects and to multiply by the number of new hyperplanes. The former number can be determined by examining the surface of each hyperplane and noting the cells into which it is divided by its intersections with the other hyperplanes; each such cell will correspond to a region through which it has passed. But how can these cells be counted? For the answer the author is indebted to Dr. Kenneth Leisenring (personal communication), who pointed out that the division of each *hyperplane* into *cells* by its intersections with other hyperplanes will be identical with the division into *regions* of the complete population space generated by one fewer subject and one fewer dimension. The reason for this identity is that each hyperplane is a legitimate subspace of the population space (that subspace from which all the variance attributable to the axis normal to the hyperplane has been extracted). All lines on its surface created by its intersections with other hyperplanes are where they ought to be if all the other subject-vectors were projected normally onto its surface and the hyperplane were treated as a population space in its own right. In this projection the two subject-vectors determining the hyperplane will be projected onto a single point, because the hyperplane was originally constructed perpendicular to the line connecting them; hence the loss of one subject as well as one dimension. So it is seen that the number $R(s, d)$ of regions or rankings created by s subjects in d-space is equal to the number present when the sth subject was added, that is, $R(s - 1, d)$, plus the number of hyperplanes added $(s - 1)$ times the number of regions which each transects, $R(s - 1, d - 1)$. The result is the recursive expression shown at the top of Table 1. This account of its derivation is rather terse because the fruit of all this labor proves to have little except theoretical interest. $R(s, d)$ quickly becomes large for moderate s and d, and so many tests are never likely to be given that the limitation on the number of different ways the subjects can be ranked becomes an important constraint on the data. It is useable directly only in the rare situation in which there are a large number of rankings of a small number of objects, as in psychophysics or in scaling experiments.

Algebraic Properties of the Permissible Rankings

The second issue is the constraint on the nature of the permissible rankings imposed by dimension, which arises in the following manner. Suppose that three subject-vectors all lie on a straight line (which does not necessarily pass through the origin) and that on this line they are in the order A, B, C. Such a configuration is illustrated in Figure 4. It is apparent that regardless of the dimensionality of the space in which these vectors occur, there are only three ways in which any linear functions of the axes can rank them: A, B, C; C, B, A; and $A = B = C$. In other words, if B is *between*

TABLE I

The Number R(s,d) of Different Ways in Which S Subjects Can Be Ranked By Linear Functions of d Factors

$$R(s,d) = R(s-1,d) + (s-1)R(s-1,d-1)$$

s	d = 2	d = 3	d = 4	d = 5
1	1	1	1	1
2	2	2	2	2
3	6 = 3!	6	6	6
4	12	24 = 4!	24	24
5	20	72	120 = 5!	120
6	30	172	480	720 = 6!
7	42	352	1,512	3,600
8	56	646	3,976	14,184
9	72	1,094	9,144	45,992
10	90	1,742	18,990	128,288
11	110	2,642	36,410	318,188
12	132	3,852	65,472	718,698
13	156	5,436	111,696	1,504,362
14	182	7,464	182,364	2,956,410
15	210	10,012	286,860	5,509,506
16	240	13,162	437,040	9,812,406
17	272	17,002	647,632	16,805,046
18	306	21,626	936,666	27,814,790
19	342	27,134	1,325,934	44,674,778
20	380	33,632	1,841,480	69,867,524
21	420	41,232	2,514,120	106,697,124
22	462	50,052	3,379,992	159,493,644
23	506	60,216	4,481,136	233,853,468
24	552	71,854	5,866,104	336,919,596
25	600	85,102	7,590,600	477,706,092
26	650	100,102	9,718,150	667,471,092
27	702	117,002	12,320,802	920,142,992
28	756	135,956	15,479,856	1,252,804,646
29	812	157,124	19,286,624	1,686,240,614
30	870	180,672	23,843,220	2,245,552,710

A and C, i.e., if B is in the interior of the interval A–C, no ranking based on a linear function of the coordinate axes is possible in which the subject B is separated from the set of subjects (A, C). Consider next the four coplanar subject-vectors in Figure 5. Note that the subject-vector D is in the interior of the triangle formed by vectors A, B, and C; consequently no linear function of the coordinate axes is possible in which the subject D is separated from the set of subjects (A, B, C). This relation may be generalized formally to an arbitrary number of dimensions in the following way: If a vector X is in the interior of a simplex formed by a vector-set S, then no linear function of the coordinate axes can generate a ranking in which the subject X is wholly separated from the set of subject S.

This relation, however true, might seem at first to have limited usefulness. What can be done with configurations like that of Figure 3, in which none of the four vectors lies in a triangle formed by the other three? The following device is proposed. Consider the subspace of dimension d containing a particular set of $d + 1$ subject-vectors in general position, that is, randomly scattered so that no two of them are coincident, no three form a line, no two pairs form parallel lines, etc. Consider next two such subspaces

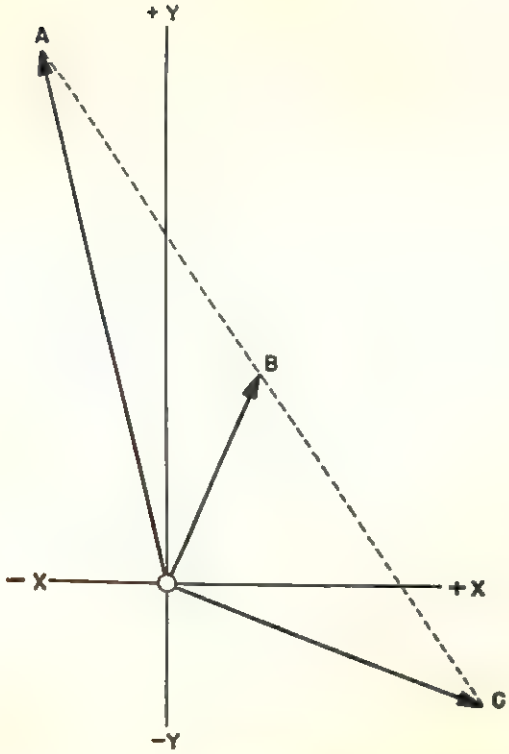


FIGURE 4

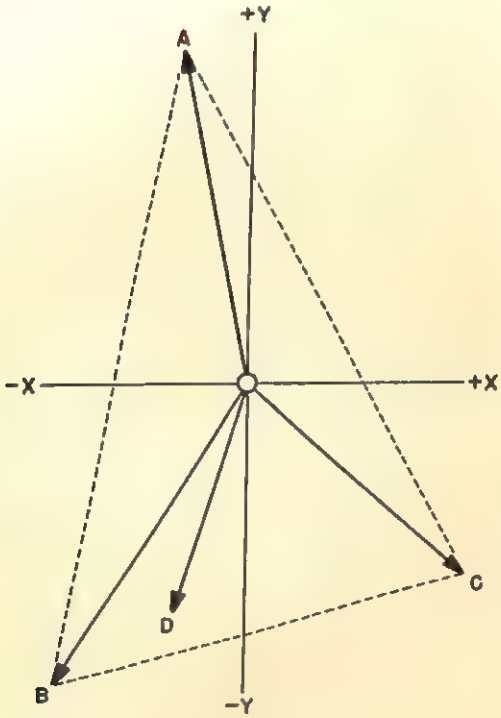


FIGURE 5

generated by two distinct subsets of subject-vectors. From familiar geometric considerations, the sum of the dimensions of two intersecting subspaces is equal to the sum of the dimensions of their union and their intersection. For example, two one-dimensional lines crossing in a plane have the whole two-dimensional plane for their union and intersect in a zero-dimensional point. In that case, one plus one equals two plus zero. Now suppose that the total number of subject-vectors in the two subsets is equal to $d + 2$, where d is the dimensionality of the population space. One of the subsets, containing n subjects, will generate an $(n - 1)$-dimensional subspace. The other, containing $d + 2 - n$ subjects, will generate a $(d + 1 - n)$-dimensional subspace. The sum of these dimensions is d. On the other hand, the dimension of their union is clearly d also, since any $d + 1$ of them are sufficient to generate the whole population space. It follows that the dimension of the intersection of the subspaces is zero; that is, they must meet in a point. Such a configuration is illustrated in Figure 6. The two subject-vectors

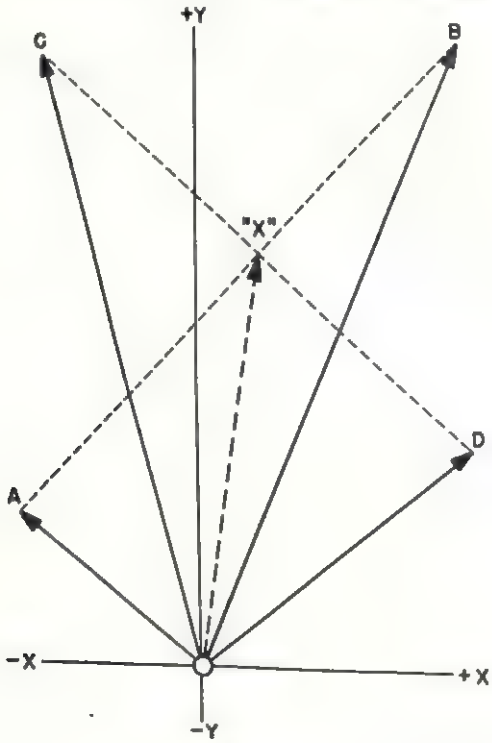


FIGURE 6

A and B generate a line $\overline{AB}$ which intersects a second line $\overline{CD}$. In this particular case, the point of intersection is between A and B on the line $\overline{AB}$ and between C and D on the line $\overline{CD}$. It is convenient to consider this point of intersection as a sort of phantom subject, X, who is a part of both subsets of subject-vectors and must obey appropriate constraints in each subset. For example, in Figure 6, no linear function of the coordinates could yield

the ranking XAB or ABX. Only AXB and BXA are possible. The same is true of the line $\overline{CD}$. Furthermore, the phantom subject must continue to obey these constraints relative to each subset when the two subsets are combined in any ranking. For example, in Figure 6, the ranking $ABCD$ is impossible, because it would call for the point of intersection X to be between A and B and at the same time to be between C and D, which is impossible. On the other hand, the order $ACBD$ is possible, because it can be written $ACXBD$, placing X in the middle of both subsets, AB and CD. The fundamental constraint which is imposed on rankings is that no ranking is possible in which this phantom subject is called upon to be in two places at once.

There are clearly several different ways in which a given set of $d + 2$ subjects can be separated into two subsets. Of these, the most important is that which arises when the point of intersection is in the interior of the simplex in both subspaces. That is, if one of the subspaces is a line, the point of intersection is between the two points which determine the line; if one of the subspaces is a plane, the point of intersection is inside the triangle formed by the three points determining the plane; and so forth. The uniqueness of the separation of the subjects into two sets which accomplishes this leads to the following proposition: If the population space is of dimension d, then given any set S of $d + 2$ subjects, there exists a unique separation of the subjects into two subsets, S^* and $S - S^*$, such that no test can rank all the subjects in S^* over all the subjects in $S - S^*$. That is, no ranking can separate the two sets entirely, because then the phantom subject, the point of intersection, could not be inside both of them at once.

The simple cases which have been discussed, for example, that of the vector which lies in the interior of the triangle formed by three other vectors in a plane, are merely special cases of this general proposition in which one of the subsets contains but a single vector.

This result is applicable, with obvious limitations, to the problem of determining the number of independent parameters of the score matrix. The score matrix for subsets of n subjects is examined. If d independent parameters are present, at least some subsets of subjects of size $d + 1$ are separated in rank by the tests in all possible ways; and, if there are enough tests, all such subsets will be separated in all possible ways. But in every subset of $d + 2$ subjects there will be one separation which never occurs in the data, no matter how the elements of each subset are permuted within the subset. This exclusion indicates that all the rankings can be accounted for by d independent parameters.

Examples from the Box Problem

Finally, examples of the exclusion property will be drawn from the score matrix of Thurstone's box problem (**3**, p. 140), to see whether, in fact, the number of independent parameters can be determined. First, consider the scores on tests 1, 2, 4, 7, 13, 14, 18, and 19, which are known to contain only

the factors X and Y. Of course when the scores are ranked the distinction among tests 1, 13, and 18 disappears at once because they are all monotonic, though non-linear, functions of X. The same is true of tests 2, 14, and 19. It is obvious that there must be more than one parameter present, because in many instances subsets of three subjects are separated in all possible ways. For example, boxes 1, 3, and 6 are variously ranked in the order 3, 6, 1 by test 4, and in the order 6, 3, 1 by test 7; no possible way of separating these three boxes into two subsets is absent. On the other hand, for every set of four boxes, there is some separation which no test imposes. For example, boxes 3 and 6 are never ranked apart from boxes 1 and 8. Box 8 is never ranked apart from boxes 1, 11, and 14. Thus two factors are sufficient to account for the score matrix. Proceeding to the analysis of all twenty tests together, it will be discovered that some sets of four subjects are separated in all possible ways. For example, boxes 1, 10, 11, and 14 are variously ranked in the order 10, 11, 14, 1 by test 17; 14, 10, 11, 1 by test 8; and 11, 14, 10, 1 by test 7. Every possible way of separating these four boxes into two subsets is represented. But in every set of five subjects a separation is discovered which does *not* occur in the data. For example, boxes 3, 5, and 7 are never separated from boxes 2 and 9. Therefore three factors are sufficient to account for all the rankings of the subjects generated by these tests.

Conclusion

The inference of factor pattern from the rankings of subjects by tests clearly has a long way to go. Both the results discussed in this paper determine dimensions in a painfully literal way. If n dimensions are needed to account for all the rankings, the technique says that there are n dimensions. But of course in any real problem there are as many dimensions as there are tests. No factor-analytic technique can hope to be useful unless some way is found to restrict the analysis to the *common* factors. In this technique, for example, it might be stipulated that *most* sets of $d + 2$ subjects shall have the exclusion property, or some such approximation. Or individual rankings might be broken down into their constituent paired comparisons, with a requirement that *nearly all* these paired comparisons shall fit the model, allowing some deviation.

Second, the determination of the number of dimensions necessary to account for the whole set of tests does not begin to settle the problem of the apportionment of the factors among the tests, that is, the questions of factor-loadings and rotation to simple structure. Thus it is still too early to say whether non-metric factor analysis of the kind proposed by Thurstone is feasible. It is the author's hope that more psychometricians will consider it worth while to find out, since the worst that can happen is that more will be learned about the underlying logic of the factor-analytic method.

REFERENCES

1. Bennett, J. F. A method for determing the dimensionality of a set of rank orders. Unpublished Ph.D. dissertation, University of Michigan, 1951.
2. Coombs, C. H. Psychological scaling without a unit of measurement. *Psychol. Rev.*, 1950, **57**, 145-158.
3. Thurstone, L. L. Multiple-factor analysis. Chicago: Univ. Chicago Press, 1947.

Manuscript received 10/24/55

Revised manuscript received 12/20/55

PSYCHOMETRIC SOCIETY
Statement of Receipts and Disbursements for Fiscal Year
Ended June 29, 1956

Receipts (Dues)

Year	Members	Student Members	
1957	1		
1956	488	58	
1955	52	14	
	541	72	$4075.00

Received with Dues for Corporation Publications	24.70
Proceeds of 1955 Joint Dinner with APA Division 5	.61
Overpayments	32.85
Miscellaneous	.08
Total Receipts	$4133.24

DISBURSEMENTS

Psychometric Corporation (90% of dues)	$3667.50
Psychometric Corporation (Publications)	24.70
Mimeographing and Printing	109.04
Postage	111.08
Secretarial Services	173.61
Addressing and Mailing	30.00
Bank Charges	3.48
1955 Special Program Expense	179.06
Refund of Overpayments	32.85
Telephone and Telegraph	2.31
Total Disbursements	$4333.63

BALANCE

Bank Balance, July 1, 1955	$1087.80
Receipts, 1955-56	4133.24
	$5221.04
Disbursements, 1955-56	4333.63
Bank Balance, June 29, 1956	$ 887.41

PSYCHOMETRIC CORPORATION
Statement of Receipts and Disbursements for Fiscal Year
Ended June 29, 1956

RECEIPTS

Subscriptions (less agency discounts)	$ 5362.30
Psychometric Society (90% of dues)	3667.50
Sale of Back Issues (less discounts)	1465.60
Sale of Monographs 5-8 (less discounts)	247.15
Royalties from U. Chi. Press (1954-55)	2.39
Interest on Savings Accounts	94.07
For Monographs 2, 4	5.50
Overpayments	11.00
Miscellaneous	7.65
Total Receipts	$10863.16

DISBURSEMENTS

Printing and Mailing Psychometrika Volume 20, No. 2, through 21, No. 1	$ 5742.15
Reprints	245.26
Stipend of Assistant Editor Volume 19, No. 3, through 20, No. 4	516.00
Secretarial Services: Editorial Office	1146.70
Secretarial Services: Business Office	303.27
Stationary and Postage	156.02
Mailing Back Issues and Monographs (12/52-3/56)	940.20
Deposited in Savings Accounts	7000.00
Refunds	38.40
Miscellaneous	42.50
Total Disbursements	$16130.50

BALANCE

Bank Balance, July 1, 1955	$ 9992.89
Receipts, 1955-56	10863.16
	$20856.05
Disbursements, 1955-56	16130.50
Bank Balance, June 29, 1956	$ 4725.55

INDEX FOR VOLUME 21

AUTHOR

Abelson, Robert P. (with S. J. Messick). The additive constant problem in multidimensional scaling. 1-15.

Angoff, William H. A note on the estimation of nonspurious correlation. 295-297.

Bennett, Joseph F. Determination of the number of independent parameters of a score matrix from the examination of rank orders. 383-393.

Bock, R. Darrell. The selection of judges for preference testing. 349-366.

Cartwright, Desmond S. A rapid non-parametric estimate of multi-judge reliability. 17-29.

Coombs, Clyde H. The scale grid: some interrelations of data models. 313-329.

Cureton, Edward E. Review of "Quinn McNemar, *Psychological Statistics* (Second Edition)." 93-94.

Cureton, Edward E. Rank-biserial correlation. 287-290.

Deuel, P. D. A nomogram for factor analysts. 291-294.

Dodd, Stuart Carter. Testing message diffusion in harmonic logistic curves. 191-205.

Feller, William. Review of "I. Blumen et al., *The Industrial Mobility of Labor as a Probability Process.*" 217.

Fishman, Joshua A. A note on Jenkins' "Improved method for tetrachoric *r*." 305.

Garner, W. R. (with W. J. McGill). The relation between information and variance analyses. 219-228.

Garside, R. F. The regression of gains upon initial scores. 67-77.

Gibson, W. A. Proportional profiles and latent structure. 135-144.

Gilbert, E. J. The matching problem. 253-266.

Green, Bert F. A method of scalogram analysis using summary statistics. 79-88.

Gulliksen, Harold. A least squares solution for paired comparisons with incomplete data. 125-134.

Gulliksen, Harold. Measurement of subjective values. 229-244.

Gulliksen, Harold. A tribute to L. L. Thurstone. 309-312.

Guttman, Louis. "Best possible" systematic estimates of communalities. 273-285.

Harris, C. W. Relationships between two systems of factor analysis. 185-190.

Horst, Paul. Optimal test length for maximum differential prediction. 51-66.

Horst, Paul (with Charlotte MacEwan). Optimum test length for maximum absolute prediction. 111-124.

Householder, A. S. Review of *Mathematical Models of Human Behavior, A Symposium*. 307-308.

Jones, Lyle V. (with W. W. Rozeboom). The validity of the successive intervals method of psychometric scaling. 165-183.

Kaiser, Henry F. Note on Carroll's analytic simple structure. 89-92.

Katz, Leo (with T. R. Wilson). The variance of the number of mutual choices in sociometry. 299-304.

Lev, Joseph. Maximizing test battery prediction when the weights are required to be non-negative. 245-252.

Lord, Frederic M. A study of speed factors in tests and academic grades. 31-50.

Luce, R. Duncan. Review of *Decision Processes*, R. M. Thrall et al. (Eds.). 96-98.

MacEwan, Charlotte (with Paul Horst). Optimum test length for maximum absolute prediction. 111-124.

McGill, William J. (with W. R. Garner). The relation between information and variance analyses. 219-228.

McHugh, Richard B. Efficient estimation and local identification in latent class analysis. 331-347.

Messick, Samuel J. An empirical evaluation of multidimensional successive intervals. 367-375.

Messick, Samuel J. (with R. P. Abelson). The additive constant problem in multidimensional scaling. 1-15.

Milholland, John E. Review of "R. L. Thorndike and Elizabeth Hagen, *Measurement and Evaluation in Psychology and Education.*" 94-95.

Rosner, Burton S. A new scaling technique for absolute judgments. 377-381.

Rozeboom, William W. (with L. V. Jones). The validity of the successive intervals method of psychometric scaling. 165-183.

Siegel, Sidney. A method for obtaining an ordered metric scale. 207-216.

Simon, Herbert A. A comparison of game theory and learning theory. 267-272.

Thorndike, Robert L. Review of "Paul E. Meehl, *Clinical and Statistical Prediction*." 218.

Torgerson, Warren S. A non-parametric test of correlation using rank orders within subgroups. 145-152.

Webster, Harold. Maximizing test validity by item selection. 153-164.

Wilson, Thurlow R. (with L. Katz). The variance of the number of mutual choices in sociometry. 299-304.